NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

64-241

AFCRL _____

STUDIES IN OPTIMUM FILTERING OF SINGLE AND MULTIPLE STOCHASTIC PROCESSES

S. M. Simpson
E. A. Robinson
R. A. Wiggins
C. I. Wunsch

Massachusetts Institute of Technology
Cambridge 39, Massachusetts

Scientific Report No. 7 of Contract
AF 19(604) 7378
S. M. Simpson, Director
June 30, 1963

Prepared for

GEOPHYSICS RESEARCH DIRECTORATE
AIR FORCE CAMBRIDGE RESEARCH CENTER
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS

WORK SPONSORED BY ADVANCED RESEARCH PROJECTS AGENCY

PROJECT VELA-UNIFORM
ARPA Order No. 180-61, Amendment 2
Project Code No. 8652, Task 865203

AFCRL 64-241

STUDIES IN OPTIMUM FILTERING OF SINGLE AND MULTIPLE STOCHASTIC PROCESSES

S. M. Simpson
E. A. Robinson
R. A. Wiggins
C. I. Wunsch

Massachusetts Institute of Technology
Cambridge 39, Massachusetts

Scientific Report No. 7 of Contract
AF 19(604) 7378
S. M. Simpson, Director
June 30, 1963

Prepared for

GEOPHYSICS RESEARCH DIRECTORATE
AIR FORCE CAMBRIDGE RESEARCH CENTER
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS

WORK SPONSORED BY ADVANCED RESEARCH PROJECTS AGENCY

PROJECT VELA-UNIFORM
ARPA Order No. 180-61, Amendment 2
Project Code No. 8652, Task 865203

ABSTRACT

This report treats the design of discrete filters for the detection of signals caused by nuclear explosions on digitized seismic recordings. The theoretical aspects of filter design are treated, together with the setting up of the necessary formulas for realizing the filters on digital computers. Specific discrete filters so treated are: (1) matched filter, (2)modified matched filter, (3)modified matched filter for a multiparameter model, (4)filter for the elimination of trend components, (5)time-invariant filter, (6)time-invariant filter in the noiseless case, (7)spike filter, (8)time-varying filter, (9)detection filter, and (10)squared magnitude devices. The normal equation forms in optimum filtering problems for the determination of the filter coefficients and the error are developed for (1)single processes, (2)multi-channel processes, and (3)multi-dimensional processes. Recursive computational schemes are presented for normal equations of Toeplitz form. For single processes the Levinson recursion for the extension of the prediction error operator and the extension of the general filter is developed, as well as the recursion to move the output origin. A corresponding development is given for multi-channel processes, as well as a development of the recursion to larger operators for the multi-dimensional processes.

The prediction problem for single stationary time series is reviewed and the least square and Kolmogoroff solutions given. Extension is then made to the multiple case, the least squares equations set up and the Wiener-Masani factorization described. Heuristic use is made of the Hilbert space property of time series. A digital computer program for performing the Wiener-Masani factorization is discussed.

Requests for additional copies by Agencies of the Department of Defense, their Contractors, and other Government Agencies should be directed to the

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

Department of Defense contractors must be established for ASTIA services or have their "need-to know" certified by the cognizant military agency of their project or contract.

ABSTRACT

This report treats the design of discrete filters for the detection of signals caused by nuclear explosions on digitized seismic recordings. The theoretical aspects of filter design are treated, together with the setting up of the necessary formulas for realizing the filters on digital computers. Specific discrete filters so treated are: (1) matched filter, (2)modified matched filter, (3)modified matched filter for a multiparameter model, (4)filter for the elimination of trend components, (5)time-invariant filter, (6)time-invariant filter in the noiseless case, (7)spike filter, (8)time-varying filter, (9)detection filter, and (10)squared magnitude devices. The normal equation forms in optimum filtering problems for the determination of the filter coefficients and the error are developed for (1)single processes, (2)multi-channel processes, and (3)multi-dimensional processes. Recursive computational schemes are presented for normal equations of Toeplitz form. For single processes the Levinson recursion for the extension of the prediction error operator and the extension of the general filter is developed, as well as the recursion to move the output origin. A corresponding development is given for multi-channel processes, as well as a development of the recursion to larger operators for the multi-dimensional processes.

The prediction problem for single stationary time series is reviewed and the least square and Kolmogoroff solutions given. Extension is then made to the multiple case, the least squares equations set up and the Wiener-Masani factorization described. Heuristic use is made of the Hilbert space property of time series. A digital computer program for performing the Wiener-Masani factorization is discussed.

TABLE OF CONTENTS

I INTRODUCTION

The detection of weak signals from nuclear explosions on seismic recordings presents an important and difficult problem, both from a practical and theoretical point of view. As a result every effort should be made to keep theory and practice coordinated and in balance with each other.

Much has been written on the design of filters for signal detection, not only in seismology but in all branches of science. The present report is unique, however, in that here the theoretical aspects of filter design are treated simultaneously with the practical setting up of the necessary formulas for realizing the filters on digital computers. Therefore by making use of the material developed in this report one can right away analyze seismic data by use of some of the most advanced filters known. The seismic data is required in digitized form, and such data is now readily available to Vela Uniform Projects from the Vela Uniform Data Center in Washington.

One of the most important contributions of this Report is the development of practical ways to design multiple-channel and multi-dimensional filters. This Report represents the first least-squares treatment of this problem. Also included is the first practical investigation of the Wiener-Masani multiple spectral factorization. Another important contribution of this Report is reflected in its completeness. Here one can find detailed treatment of many important types of filters, some presented for the first time from the digital point of view.

NOTATION CONVENTIONS

In order to preserve a general consistancy in the notation used in this report, we have adopted the following conventions.

1) Division of the alphabet

Specific use of the letters of the alphabet are assigned in the body of the report. However, a general division is

A, B, C	Transients (wavelets, operators, filters, etc.)
X, Y, Z	(Stationary) General Processes

2) Use of Subscripts and Superscripts

A discrete multi-dimensional process is designated by

$$ {}^{N}X_{t}^{m,n,\ell} $$

where t is the time index

m, n, ℓ are space indices

and N is open to any particular interpretation.

The dimensionality of the process is given by the total number of super- and sub-scripts to the right of the letter. Thus, the example above has 4 dimensions.

The order of a process is given by the number of equivalent one-dimensional processes in a multiple process. Thus, it is given by the product of the maximum values assumed by the superscripts.

3) Upper and Lower case; Script and Non-script Letters

Unless otherwise defined, the following conventions for upper and lower case, and script and non-script letters will be used:

	Single Process	Multiple Process
Time Domain	x_t Y_t r_t	$x_t^{n,m}$ $y_t^{n,m}$ $r_t^{n,m}$
Frequency Domain	$X(\omega)$, $Y(\omega)$	$\mathcal{X}^{n,m}(\omega)$, $\mathcal{Y}^{n,m}(\omega)$

2. Discrete Filters for Digital Data

DISCRETE FILTERS FOR DIGITAL DATA

1. Matched filter

Assumptions:

(a) The signal s_t has a known fixed shape.
(b) The noise n_t is a normal random process.
(c) The mean value $E\{n_t\}$ is known to be equal to zero.
(d) The covariance $\phi_{tr} = E\{n_t n_r\}$ of the noise is known.
(e) Time t is a discrete, integer-valued parameter.
(f) The observed random process is x_t.
(g) The time period under observation is $t = 1,2,\ldots,N$

Problem:

We wish to test whether hypothesis H_o is true or hypothesis H_1 is true, where

$$H_o: \quad x_t = n_t$$
$$H_1: \quad x_t = s_t + n_t$$

Matrix notation:

$x = (x_1, x_2, \ldots, x_N)$; 1xN row vector
$n = (n_1, n_2, \ldots, n_N)$; 1xN row vector
$s = (s_1, s_2, \ldots, s_N)$; 1xN row vector

$$\phi = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \cdots & & \cdots & \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} \end{bmatrix} = [\phi_{tr}]$$: NxN covariance matrix of the noise (assumed to be non-singular)

$\det \phi$: determinant of the covariance matrix ϕ

$\mu = [\mu_{tn}] = \phi^{-1}$: inverse of the covariance matrix ϕ

A prime indicates matrix transpose.

Probability density

Under hypothesis H_0, x is normally distributed with mean zero and covariance matrix ϕ ; that is, the probability density $f_0(x)$ is

$$f_0(x) = \frac{1}{(2\pi)^{N/2} (\det \phi)^{1/2}} \exp\left[-\tfrac{1}{2} x \mu x'\right]$$

Under hypothesis H_1, x is normally distributed with mean s and covariance matrix ϕ ; that is, the probability density $f_1(x)$ is

$$f_1(x) = \frac{1}{(2\pi)^{N/2} (\det \phi)^{1/2}} \exp\left[-\tfrac{1}{2} (x-s) \mu (x-s)'\right]$$

Likelihood ratio

The likelihood ratio $\Lambda(x)$ is formed by taking the quotient of these two density functions, that is

$$\Lambda(x) = \frac{f_1(x)}{f_0(x)} = \exp\left[-\tfrac{1}{2}\left\{(x-s)\mu(x-s)' - x\mu x'\right\}\right]$$

$$= \exp\left[-\tfrac{1}{2}\left\{x\mu x' - x\mu s' - s\mu x' + s\mu s' - x\mu x'\right\}\right]$$

$$= \exp\left[-\tfrac{1}{2}\left\{-2 x\mu s' + s\mu s'\right\}\right]$$

(since $x\mu s' = s\mu x'$). The observer will choose hypothesis H_o when

$$\Lambda(x) < \Lambda_o$$

where Λ_o is a constant determined by the decision criterion used. Setting $\Lambda(x) = \Lambda_o$, and taking logarithms, we obtain

$$-\frac{1}{2}\{-2x\mu s' + s\mu s'\} = \log \Lambda_o.$$

Thus the inequality $\Lambda(x) < \Lambda_o$ is the same as the inequality

$$-\frac{1}{2}\{-2x\mu s' + s\mu s'\} < \log \Lambda_o$$

or, what is the same thing,

$$x\mu s' < \log \Lambda_o + \frac{1}{2} s\mu s'$$

Decision rule:

Setting $G = \log \Lambda_o + \frac{1}{2} s\mu s'$, we have the following decision rule:

> Choose H_o (that is, say that the signal is not present) if
>
> $$x\mu s' < G.$$
>
> Choose H_1 (that is, say that the signal is present) if
>
> $$x\mu s' > G.$$

That is, the decision is based on the test statistic

$$x\mu s' = \sum_{t=1}^{N} \sum_{r=1}^{N} x_t \mu_{tr} s_r$$

computed from the observed process

$$x = (x_1, x_2, \cdots, x_N)$$

and the known signal

$$S = (s_1, s_2, \cdots, s_N)$$

Hence we may say that the observed process x is compared with, or matched to, the signal s. For this reason, the filter which performs the computation $x\mu s'$ is called a matched filter.

Distribution of the test statistic:

Since the test statistic $x\mu s'$ is a linear combination of normal random variables x, it follows that $x\mu s'$ itself is normally distributed, with mean and variance as follows:

Under H_o (i.e. x = n):

$$E\{x\mu s'\} = E\{x\}\mu s' = 0 \quad \text{since } E\{x\} = 0.$$

$$\text{var}\{x\mu s'\} = E\{(x\mu s')^2\} = E\{(x\mu s')' x\mu s'\}$$

$$= E\{s\mu' x' x \mu s'\}$$

$$= s\mu' E\{x'x\}\mu s' = s\mu'\phi\mu s'$$

$$= s\mu s'$$

since $E\{x'x\} = \phi = E\{n'n\}$,

$\phi\mu = I$ = the identity matrix,

$\mu' = \mu$.

Under H_1 (i.e. $x = s + n$):

$$E\{x\mu_s'\} = E\{x\}\mu_s' = s\mu_s' \quad \text{since } E\{x\} = s.$$

$\text{var}\{x\mu_s'\} = s\mu_s'$ (the same as the variance under H_o, because the variance of a random variable doesn't depend upon its mean).

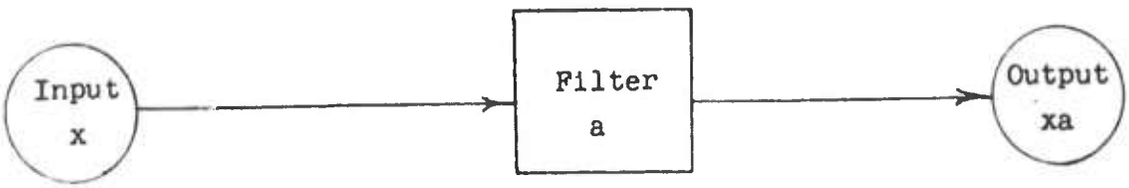
Relaxation of the normality assumption

Let us now drop assumption (b), that is, we no longer assume that n_t is a normal random process, but instead, we assume that: (b_1) n_t is a random process with unknown distribution.

We now consider a filter with coefficients given by

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix} \quad : \quad \text{Nx1 column vector}$$

The input to this filter is the observed random process x and the output is xa, that is:



The mean-square output is defined to be

$$E\{(xa)^2\} = E\{(xa)'xa\} = E\{a'x'x\,a\}$$

$$= a'\,E\{x'x\}\,a.$$

Under hypothesis H_o (x = n, or no signal present), the mean-square output is

$$E\{(xa)^2\} = a' E\{n'n\} a = a' \phi a .$$

Under hypothesis H_1 (x = s + n, or signal present), the mean-square output is

$$E\{(xa)^2\} = a' E\{(s+n)'(s+n)\} a$$

$$= a' E\{s's\}a + a' E\{s'n\}a + a' E\{n's\}a + a' E\{n'n\}a$$

$$= a's'sa + a'\phi a$$

since

$$E\{n\} = 0 , \quad E\{n'\} = 0 , \quad E\{n'n\} = \phi$$

We wish to determine a such that the ratio

$$\frac{\text{Mean-square output under } H_1}{\text{Mean-square output under } H_o} = \frac{a's'sa + a'\phi a}{a'\phi a} = 1 + \frac{a's'sa}{a'\phi a}$$

is a maximum, or equivantly such that the ratio

$$\lambda = \frac{a's'sa}{a'\phi a}$$

is a maximum.

Hence we require

$$\frac{\partial \lambda}{\partial a} = \frac{(a'\phi a)\, 2s'sa - (2\phi a)\, a's'sa}{(a'\phi a)^2} = 0$$

or

$$2S'Sa - 2\phi a\left(\frac{a'S'Sa}{a'\phi a}\right) = 0$$

or

$$S'Sa - \phi a\lambda = 0$$

The solution is

$$a = \mu S'$$

The output of the filter is

$$\gamma a = \gamma \mu S'$$

which is the same as the test statistic obtained under the assumption that n was normally distributed.

2. Modified matched filter

Assumptions

We now wish to modify assumption (a) given in section 1 as follows:

(a_1) The signal s_t is given by

$$s_t = c f_t$$

where c is an unknown constant and f_t is a known fixed function.

Problem

We wish to test whether hypothesis H_o or hypothesis H_1 is true, where

$$H_o: \ \gamma_t = n_t \qquad (c = 0)$$

$$H_1: \ \gamma_t = c f_t + n_t \qquad (c \neq 0)$$

Matrix notation

$f = (f_1, f_2, \ldots, f_N)$: 1xN row vector

$s = cf$

Likelihood ratio

The likelihood ratio is

$$\Lambda(x) = \exp\left[-\frac{1}{2}\{-2x\mu(cf)' + cf\mu(cf)'\}\right]$$

$$= \exp\left[c\, x\mu f' - \frac{c^2}{2} f\mu f'\right]$$

The observer will choose H_0 when the likelihood ratio Λ satisfies $\Lambda < \Lambda_0$ for some fixed threshold Λ_0. Because c is unknown, we consider

$$\max_{-\infty<c<\infty} \Lambda(x) < \Lambda_0$$

instead of $\Lambda(x) < \Lambda_0$. The maximum occurs when the negative of the exponent of $\Lambda(x)$ is a minimum, that is, when

$$J = -c\, x\mu f' + \frac{c^2}{2} f\mu f' = \text{minimum}$$

We have

$$\frac{\partial J}{\partial c} = 0 :- x\mu f' + \hat{c}\, f\mu f' = 0$$

or

$$\boxed{\hat{c} = \frac{x\mu f'}{f\mu f'}}$$

Hence

$$\max_{-\infty<c<\infty} \Lambda(x) = \exp\left[\hat{c}\, x\mu f' - \frac{\hat{c}^2}{2} f\mu f'\right]$$

$$= \exp\left[\frac{x\mu f'}{f\mu f'} x\mu f' - \frac{(x\mu f')^2}{2(f\mu f')^2} f\mu f'\right]$$

$$= \exp\left[\frac{1}{2}\frac{(x\mu f')^2}{f\mu f'}\right]$$

Taking logarithms, we have

$$\frac{1}{2}\frac{(x\mu f')^2}{f\mu f'} < \log \Lambda_0 .$$

Decision rule

Letting G = 2 log Λ_0, we obtain the decision rule:

Choose H_0 (that is, say that the signal is not present) if

$$\frac{(x\mu f')^2}{f\mu f'} < G.$$

Choose H_1 (that is, say that the signal is present) if

$$\frac{(x\mu f')^2}{f\mu f'} > G.$$

That is, the decision is based on the <u>test statistic</u>

$$\frac{(x\mu f')^2}{f\mu f'} = \frac{\left(\sum_{t=1}^{N}\sum_{r=1}^{N} x_t \mu_{tr} f_r\right)^2}{\sum_{t=1}^{N}\sum_{r=1}^{N} f_t \mu_{tr} f_r}$$

computed from the observed process

$$x = (x_1, x_2, \cdots, x_N)$$

and the known

$$f = (f_1, f_2, \cdots, f_N).$$

We see that the test statistic is a quadratic function of the observations $x = (x_1, x_2, \ldots, x_N)$.

<u>Maximum likelihood estimates</u>

Let us now find the maximum likelihood estimates of c

and s. We recall that the density function of x, under the normality assumption, is

$$f(x) = \frac{1}{(2\pi)^{N/2}(\det\phi)^{1/2}} \exp\left[-\tfrac{1}{2}(x-s)\mu\,(x-s)'\right] \quad \text{where } s = cf.$$

The maximum-likelihood estimate of s, denoted by $\hat{s}$, is that value of $s = cf$ for which $f(x)$ is a maximum. The maximum occurs when the negative of the exponent of $f(x)$ is a minimum that is, when

$$P = (x-s)\mu\,(x-s)' = \text{minimum},$$

subject to the constraint that $s = cf$. We shall now use the method of Lagrange multipliers to solve this constrained minimization problem. We therefore introduce the undetermined multipliers

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{pmatrix} \quad : \quad \text{Nx1 column vector.}$$

We now wish to minimize

$$J = (x-s)\mu(x-s)' + (s-cf)\lambda + \lambda'(s-cf)'$$

with respect to s, λ, and c. We thus obtain:

$$\frac{\partial J}{\partial s} = 0 \quad : \quad -(x-\hat{s})\mu + \hat{\lambda}' = 0$$

$$\frac{\partial J}{\partial \lambda} = 0 \quad : \quad \hat{s} - \hat{c}f = 0$$

$$\frac{\partial J}{\partial c} = 0 \qquad f'\hat{\lambda}'f = 0 .$$

Solving these equations, we have

$$\hat{\lambda}' = (x - \hat{s})\mu = (x - \hat{c}f)\mu$$

$$\hat{\lambda}' f' = (x - \hat{c}f)\mu f' = 0 \quad ,$$

or

$$x\mu f' - \hat{c} f \mu f' = 0 \, .$$

Hence

$$\boxed{\hat{c} = \frac{x\mu f'}{f\mu f'} \quad , \quad \hat{s} = \hat{c}f = \frac{x\mu f'}{f\mu f'} f}$$

where $\hat{s}$ is the maximum-likelihood estimate of s.

Let us now find the maximum-likelihood estimate of c. By the above reasoning, it is that value of c obtained by requiring

$$P = (x-s)\mu(x-s)' = (x-cf)\mu(x-cf)' = \text{minimum}$$

We thus have

$$\frac{\partial P}{\partial c} = 0 \; : \; -(x - \hat{c}f)\mu f' = 0 \quad , \text{or} \quad \hat{c} = \frac{x\mu f'}{f\mu f'} \, ,$$

which is the same $\hat{c}$ as found above, as we would expect.

Miscellaneous notes

We recall that

$$x = (x_1, x_2, \cdots, x_N)$$

$$f = (f_1, f_2, \cdots, f_N)$$

$$\phi = \begin{pmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \cdots & & & \cdots \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} \end{pmatrix} \qquad \text{(non-singular matrix)}$$

$$\mu = \phi^{-1}$$

$$\hat{C} = \frac{x \mu f'}{f \mu f'} .$$

Now it may be shown that the quadratic form $x\mu f'$ is (where $|\ |$ indicates determinant):

$$x\mu f' = - \begin{vmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} & f_1 \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} & f_2 \\ \cdots & & & \cdots & \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} & f_N \\ x_1 & x_2 & \cdots & x_N & 0 \end{vmatrix} \div \begin{vmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \cdots & & & \cdots \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} \end{vmatrix}$$

Similarly

$$f\mu f' = - \begin{vmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} & f_1 \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} & f_2 \\ \cdots & & & \cdots & \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} & f_N \\ f_1 & f_2 & \cdots & f_N & 0 \end{vmatrix} \div \begin{vmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2N} \\ \cdots & & & \cdots \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} \end{vmatrix}$$

Hence

$$\hat{C} = \frac{x\mu f'}{f\mu f'} = \begin{vmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} & f_1 \\ \cdots & & & \cdots & \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} & f_N \\ x_1 & x_2 & \cdots & x_N & 0 \end{vmatrix} \div \begin{vmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1N} & f_1 \\ \cdots & & & \cdots & \\ \phi_{N1} & \phi_{N2} & \cdots & \phi_{NN} & f_N \\ f_1 & f_2 & \cdots & f_N & 0 \end{vmatrix}$$

3. Modified matched filter for a multiparameter model

Assumptions

Instead of assumption (a_1) given at the beginning of Section 2, let us introduce assumption

(a_2) The signal s_t is given by

$$s_t = \sum_{i=1}^{p} c_i f_{it}$$

where $c_1, c_2, \ldots, c_p$ are unknown constants and $f_{1t}, f_{2t}, \ldots, f_{pt}$ are known fixed functions.

Problem

We wish to test:

$$H_0 : x_t = n_t \qquad (c_1 = 0, c_2 = 0, \cdots, c_p = 0)$$

$$H_1 : x_t = \sum_{i=1}^{p} c_i f_{it} + n_t \qquad (\text{some or all of the } c_1, c_2, \cdots, c_p \neq 0)$$

Matrix notation:

$$c = (c_1, c_2, \cdots, c_p) \quad : \quad 1 \times p \text{ row vector}$$

$$f = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1N} \\ f_{21} & f_{22} & \cdots & f_{2N} \\ \cdots & & & \cdots \\ f_{p1} & f_{p2} & \cdots & f_{pN} \end{pmatrix} \quad : \quad p \times N \text{ matrix}$$

$$s = cf$$

Likelihood ratio

The likelihood ratio is

$$\Lambda(x) = \exp\left[x \mu f' c' - \frac{1}{2} c f \mu f' c' \right]$$

The maximum of Λ (x) with respect to all values of c occurs when

$$x_\mu f' - \hat{c} f_\mu f' = 0, \quad \text{or}$$

$$\boxed{\hat{c} = x_\mu f' (f_\mu f')^{-1}}$$

This maximum is

$$\max_{-\infty < c < \infty} \Lambda(x) = \exp\left[x_\mu f' \hat{c}' - \tfrac{1}{2} \hat{c} f_\mu f' \hat{c}' \right]$$

$$= \exp\left[x_\mu f' (f_\mu f')^{-1} f_\mu x' - \tfrac{1}{2} x_\mu f' (f_\mu f')^{-1} f_\mu f' (f_\mu f')^{-1} f_\mu x' \right]$$

$$= \exp\left[\tfrac{1}{2} x_\mu f' (f_\mu f')^{-1} f_\mu x' \right]$$

where we have used $\left[(f_\mu f')^{-1}\right]' = (f_\mu f')^{-1}$ and $\mu' = \mu$.

Decision rule:

As before, we let $G = 2 \log \Lambda_0$. We thus obtain the decision rule:

Choose H_0 if $(x_\mu f')(f_\mu f')^{-1} f_\mu x' < G$

Choose H_1 if $(x_\mu f')(f_\mu f')^{-1} f_\mu x' > G$

Maximum-likelihood estimates

As before, we find the maximum-likelihood estimates of c and s are

$$\hat{c} = (x_\mu f')(f_\mu f')^{-1}$$
$$\hat{s} = \hat{c} f = (x_\mu f')(f_\mu f')^{-1} f .$$

Let us now find the expected value of $\hat{c}$. It is

$$E\{\hat{c}\} = E\{x\mu f'(f\mu f')^{-1}\} = E\{x\}\mu f'(f\mu f')^{-1}$$

$$= cf\mu f'(f\mu f')^{-1} = c .$$

Hence $E\{\hat{c}\} = c$ so $\hat{c}$ is an unbiased estimate of c. Also

$$E\{\hat{s}\} = E\{\hat{c}f\} = cf = s$$

so $\hat{s}$ is an unbiased estimate of s.

The covariance matrix of $\hat{c}$ is

$$\Psi = E\{(\hat{c}-c)'(\hat{c}-c)\} = E\{[(x-cf)\mu f'(f\mu f')^{-1}]'[(x-cf)\mu f'(f\mu f')^{-1}]\}.$$

But under both hypothesis H_0 and hypothesis H_1, we have

$$x - cf = n .$$

Hence

$$\Psi = E\{[n\mu f'(f\mu f')^{-1}]'[n\mu f'(f\mu f')^{-1}]\}$$

$$= E\{(f\mu f')^{-1} f\mu n'n\mu f'(f\mu f')^{-1}\}$$

$$= (f\mu f')^{-1} f\mu E\{n'n\}\mu f'(f\mu f')^{-1}\}$$

$$= (f\mu f')^{-1} f\mu \phi\mu f'(f\mu f')^{-1}$$

$$= (f\mu f')^{-1} f\mu f'(f\mu f')^{-1} \qquad (\text{since } \phi\mu = I)$$

$$= (f\mu f')^{-1}.$$

Hence $\hat{c}$ has mean c and covariance matrix $(f\mu f')^{-1}$.

The covariance matrix of $\hat{s}$ is

$$E\{(\hat{s}-s)'(\hat{s}-s)\} = E\{[(x-cf)\mu f'(f\mu f')^{-1}f]'[(x-cf)\mu f'(f\mu f')^{-1}f]\}$$

$$= f'(f\mu f')^{-1} f\mu \; E\{x'x\} \mu f'(f\mu f')^{-1} f$$

$$= f'(f\mu f')^{-1} f\mu \phi \mu f'(f\mu f')^{-1} f$$

$$= f'(f\mu f')^{-1} f$$

Hence $\hat{s} = \hat{c}f$ has mean $s = cf$ and covariance matrix $f'(f\mu f')^{-1}f$.

Miscellaneous notes

We note that the maximum likelihood estimates $\hat{c}$ and $\hat{s}$ are independent of the scale of the covariance matrix. That is, suppose

$$\phi = \sigma^2\psi \quad \text{and} \quad \mu = \phi^{-1} = \sigma^{-2}\psi^{-1}$$ where ψ is known but the scale factor σ^2 is unknown. Then

$$\hat{c} = x\psi^{-1}f'(f\psi^{-1}f')^{-1}, \quad \hat{s} = \hat{c}f,$$

independent of σ^2. Thus assumption (d) may be so relaxed.

4. Elimination of the trend components

Assumptions

As a preliminary step in the analysis of random processes, one might try various methods to eliminate as well as possible any trend components. One approach to this problem is to make use of the models discussed in the foregoing sections, with the

following changes in terminology.

(1) s_t, instead of being called the signal, is now called the trend component.

(2) n_t, instead of being called the noise, is now called the trend-free component.

(3) As before, the observed random process is x_t.

We shall make assumptions (a_2), stated in Section 3, and (b), (c), (d), (e), (f), (g), stated in Section 1.

Problem

Given that $x_t = s_t + n_t$, estimate the trend component s_t.

Maximum likelihood estimates

As a solution to the problem, the maximum likelihood estimates of c and s may be used. They were given at the end of Section 3, and we recall that they are:

$$\hat{c} = (x\mu f')\,(f\mu f')^{-1}$$

$$\hat{s} = \hat{c}f = (x\mu f')(f\mu f')^{-1}f$$

where ϕ is the covariance matrix of the trend-free component n_t, and where

$$\mu = \phi^{-1}$$

Also we recall that we may relax assumption (d) to

(d_1) The covariance $\phi_{tr} = E\{n_t n_r\}$ is equal to

$$\phi_{tr} = \sigma^2 \psi_{tr}$$

where ψ_{tr} is known and σ^2 is an unknown scale factor. Then we let $\phi = \sigma^2\psi$, and hence $\mu = \sigma^{-2}\psi^{-1}$. Substituting $\mu = \sigma^{-2}\psi^{-1}$

into the expressions for $\hat{c}$ and $\hat{s}$ we obtain

$$\hat{c} = x\psi^{-1}c'(f\psi^{-1}f')^{-1}$$
$$\hat{s} = x\psi^{-1}f'(f\psi^{-1}f')^{-1}f$$

which shows that the maximum likelihood estimates are independent of the unknown scale factor, σ^2 .

Relaxation of the normality assumption

Instead of (b) we now assume:

(b_1) n_t is a random process with an unknown distribution.

Least-squares estimate

We have

$$x = cf + n$$

and we wish to estimate c. The least-squares estimate $\check{c}$ is that value of c such that the sum of squared errors

$$J = (x - cf)(x - cf)' = \text{minimum.}$$

We have

$$\frac{\partial J}{\partial c} = 0 \quad : \quad (x - \check{c}f)(-f') = 0,$$

or

$$\check{c}ff' = xf'$$

Hence the least-squares estimate is

$$\boxed{\check{c} = xf'(ff')^{-1}}$$

To compute $\check{c}$ we see that we do not need assumption (d); that is, we do not need to know the covariance matrix ϕ of n_t in order to obtain the least-squares estimate $\check{c}$. This fact makes the least-squares estimate extremely useful in practice.

The least-squares estimate is unbiased since

$$E\{\check{c}\} = E\{x f'(ff')^{-1}\} = E\{x\} f'(ff')^{-1}$$

(and using $E\{x\} = cf$)

$$E\{\check{c}\} = cff'(ff')^{-1} = c.$$

The covariance matrix of $\check{c}$ is defined to be

$$E\{(\check{c}-c)'(\check{c}-c)\}$$

Since

$$\begin{aligned}\check{c} &= x f'(ff')^{-1} = (n+s) f'(ff')^{-1} \\ &= (n+cf)(f'(ff')^{-1} \\ &= n f'(ff')^{-1} + cff'(ff')^{-1} \\ &= n f'(ff')^{-1} + c\end{aligned}$$

we have

$$\check{c} - c = n f'(ff')^{-1}$$

Hence the covariance matrix of $\check{c}$ is

$$E\{(\check{c}-c)'(\check{c}-c)\} = E\{[n f'(ff')^{-1}]'[n f'(ff')^{-1}]\}$$

$$= E\{(ff')^{-1} f n'n f'(ff')^{-1}\}$$

$$= (ff')^{-1} f E\{n'n\} f'(ff')^{-1}$$ which is

$$\boxed{E\{(\check{c}-c)'(\check{c}-c)\} = (ff')^{-1} f \phi f'(ff')^{-1}}$$

Best unbiased linear estimate

We wish to find an estimate $\hat{c}$ which

(1) is linear in the observations, i.e.

$$\hat{c} = x\,b$$

where

$$b = \begin{pmatrix} b_{11} & b_{12} \cdots b_{1p} \\ b_{21} & b_{22} \quad b_{2p} \\ \\ b_{N1} & b_{N2} \cdots b_{Np} \end{pmatrix}$$

(2) is an unbiased estimate, i.e.

$$E\{\hat{c}\} = c$$

(3) is best in the sense that the covariance matrix of $\hat{c}$ is "less than" the covariance matrix of any other linear unbiased estimate $\tilde{c}$, i.e.

$$E\{(\hat{c}-c)'(\hat{c}-c)\} < E\{(\tilde{c}-c)'(\tilde{c}-c)\}$$

(More precisely, the "less than" sign $<$ as used here means that

$$E\{(\tilde{c}-c)'(\tilde{c}-c)\} - E\{(\hat{c}-c)'(\hat{c}-c)\}$$

is a positive definite matrix.)

Using (1) we have

$$\hat{c} = x\,b = (s+n)\,b = (cf+n)\,b$$
$$= cfb + nb$$

Using (2) we have

$$c = E\{\hat{c}\} = E\{cfb\} + E\{nb\}$$

$$= cfb + E\{n\}b$$

$$= cfb \qquad \text{since } E\{n\}=0$$

Thus we have the constraint

$$fb = I = p \times p \text{ identity matrix.}$$

Also we have $\hat{c}-c=nb$ so the covariance matrix of $\hat{c}$ is

$$E\{(\hat{c}-c)'(\hat{c}-c)\} = E\{(nb)'(nb)\}$$

$$= E\{b'n'nb\} = b'E\{n'n\}b = b'\phi b.$$

Hence b may be determined as follows:

> Minimize the pxp matrix $b'\phi b$
> subject to the constraint $fb = I$.

We may use the method of Lagrange multipliers. Introduce the pxp matrix λ as an undetermined multiplier. We then wish to minimize

$$J = b'\phi b + 2\lambda(fb - I)$$

with respect to b and λ. We have

$$\frac{\partial J}{\partial b} = 0 : \quad b'\phi + \lambda f = 0$$

$$\frac{\partial J}{\partial \lambda} = 0 : \quad fb = I$$

Solving these equations for b and λ we have

$$b' = -\lambda f \phi^{-1}$$

Thus $b = -\phi^{-1}f'\lambda'$.

Also $fb = I$ gives $-f\phi^{-1}f'\lambda' = I$,

or

$$\lambda' = -(f\phi^{-1}f')^{-1}$$

Hence

$$b = -\phi^{-1}f'\lambda' = \phi^{-1}f'(f\phi^{-1}f')^{-1}.$$

We recall the notation for ϕ^{-1} was

$$\mu = \phi^{-1}$$

Hence

$$b = \mu f'(f\mu f')^{-1}$$

Thus the best unbiased linear estimate $\hat{c}$ of c is

$$\boxed{\hat{c} = xb = x\mu f'(f\mu f')^{-1}}$$

which, we see, is the same as the maximum-likelihood estimate obtained under the normality assumption (b). The covariance matrix of $\hat{c}$ is

$$E\{(\hat{c}-c)'(\hat{c}-c)\} = b'\phi b = (f\mu f')^{-1}f\mu\phi\mu f'(f\mu f')^{-1}$$

which is

$$\boxed{E\{(\hat{c}-c)'(\hat{c}-c)\} = (f\mu f')^{-1}}$$

5. Time-invariant filter

5.1 Assumptions:

(a) The signal s_t has a known fixed shape.

(b) The noise n_t is a random process with an unknown distribution.

(c) The mean value $E\{n_t\}$ of the noise is known to be equal to zero.

(d) The covariance $\phi_{tr} = E\{n_t n_r\}$ of the noise is known.

(e) Time t is a discrete, integer-valued parameter.

(f) The observed random process is $x_t = s_t + n_t$

(g) The random process x_t is observed for t = 1, 2, ..., N

(h) The observed random process $x_1, x_2, \ldots, x_N$ is passed into a time-invariant linear filter with coefficients $\beta_0, \beta_1, \ldots \beta_M$ (to be determined).

(i) The actual output of the filter is

$$y_t = \sum_{s=0}^{M} \beta_s x_{t-s} \qquad (t = 0, 1, 2, \cdots, M+N)$$

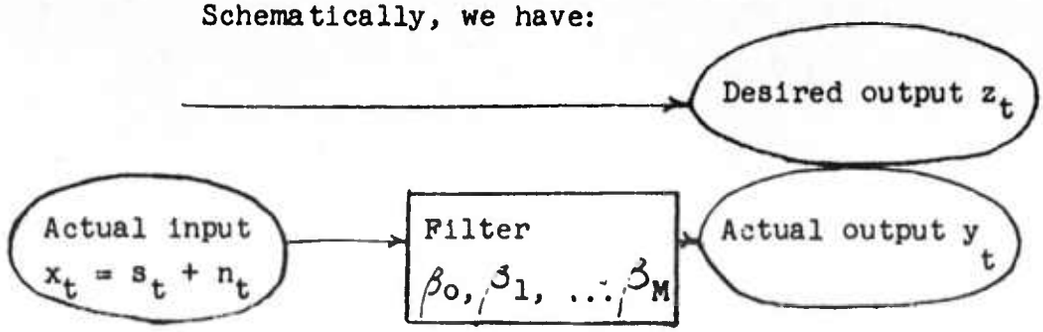
(j) The desired output of the filter is Z_t (t = 0, 1, 2, ... M+N), where Z_t is a known fixed function.

5.2 Problem:

We wish to determine those values of the coefficients $\beta_0, \beta_1, \ldots, \beta_M$ such that the mean of the sum of squared-errors between the desired output Z_t and the actual output y_t is a minimum, that is, such that

$$E\left\{\sum_{t=0}^{M+N} (Z_t - y_t)^2\right\} = \text{minimum}$$

Schematically, we have:



5.3 Matrix notation:

We define the $(N + M + 1) \times (M + 1)$ matrices:

$$X = \begin{bmatrix} x_0 & 0 & \cdots & 0 \\ x_1 & x_0 & & 0 \\ \vdots & & & \\ x_N & x_{N-1} & & \vdots \\ 0 & x_N & & \\ \vdots & & & \\ 0 & 0 & & x_{N-1} \\ 0 & 0 & & x_N \end{bmatrix}$$

$$S = \begin{bmatrix} s_0 & 0 & \cdots & 0 \\ s_1 & s_0 & & 0 \\ \vdots & & & \\ s_N & s_{N-1} & & \vdots \\ 0 & s_N & & \\ \vdots & & & \\ 0 & 0 & & s_{N-1} \\ 0 & 0 & & s_N \end{bmatrix}$$

$$N = \begin{bmatrix} n_0 & 0 & \cdots & 0 \\ n_1 & n_0 & & 0 \\ \vdots & & & \\ n_N & n_{N-1} & & \vdots \\ 0 & n_N & & \\ \vdots & & & \\ 0 & 0 & & n_{N-1} \\ 0 & 0 & & n_N \end{bmatrix}$$

Thus

$$X = S + N$$

Let y, z, and e be the (N + M + 1) x 1 column vectors

$$y = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{M+N} \end{bmatrix}, \quad z = \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{M+N} \end{bmatrix}, \quad e = z - y = \begin{bmatrix} e_0 \\ e_1 \\ \vdots \\ e_{M+N} \end{bmatrix}$$

and let β be the (M + 1) x 1 column vector

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_M \end{bmatrix}$$

Thus

$$y = X\beta$$

5.4 Determination of β

We wish to find that value $\check{\beta}$ of β that minimizes

$$J = E\left\{ \sum_{t=0}^{M+N} (z_t - y_t)^2 \right\} = E\left\{ (z-y)'(z-y) \right\} = E\left\{ (z-X\beta)'(z-X\beta) \right\}.$$

Now

$$J = E\left\{ (z-X\beta)'(z-X\beta) \right\} = E\left\{ z'z - z'X\beta - \beta'X'z + \beta'X'X\beta \right\}$$

Because $X = S + N$ and $E\{N\} = 0$, $E\{N'\} = 0$,

we have

$$J = E\{ z'z - 2z'X\beta + \beta'X'X\beta \}$$
$$= E\{ z'z - 2z'(S+N)\beta + \beta'(S+N)'(S+N)\beta \}$$
$$= E\{ z'z - 2z'S\beta + \beta'S'S\beta + \beta'N'N\beta \}$$
$$= z'z - 2z'S\beta + \beta'S'S\beta + \beta' E\{N'N\}\beta$$

Hence

$$\frac{\partial J}{\partial \beta} = 0 \quad : \quad -2z'S + 2\check{\beta}'S'S + 2\check{\beta}' E\{N'N\} = 0$$

or

$$S'z = S'S\check{\beta} + E\{N'N\}\check{\beta}$$

or

$$\boxed{\check{\beta} = \left[S'S + E\{N'N\}\right]^{-1} S'z}$$

6. Time-invariant filter in the noiseless case

6.1 Assumptions

The assumptions are the same as those given in Subsection 5.1, except that now it is assumed that there is no noise (i.e. $n_t \equiv 0$). In other words, assumptions (b), (c), (d) are omitted, and (f) becomes

(f_1) The observed process x_t is the fixed function s_t, i.e. $x_t \equiv s_t$.

6.2 Problem

We wish to find that value $\check{\beta}$ of β for which the sum of squared errors between the desired output z and the actual output y is a minimum, that is, such that

$$\sum_{t=0}^{M+N} (z_t - y_t)^2 = (z-y)'(z-y) = \text{minimum}.$$

6.3 Determination of β

From the result of the last Section (Subsection 5.4) the desired $\check{\beta}$ is

$$\check{\beta} = [S'S]^{-1} S'z .$$

7. Spike filter

7.1 Problem

The spike filter is a specialization of the time-invariant filter. The spike filter is designed such that with the signal as input it will produce little or no output while the signal is entering the filter, a large positive spike when the signal has fully entered the filter, and little or no output thereafter. The spike filter is also designed to have little output when noise is its only input.

Suppose we want the spike to occur at time $t = t_o$. Then we let the desired output z_t be

$$z_t = \begin{cases} 1 & \text{when } t = t_o \\ 0 & \text{when } t \neq t_o \end{cases}$$

or in matrix notation

$$z = \begin{bmatrix} 0 \\ ? \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where the one occurs in the position of z_{t_0}.

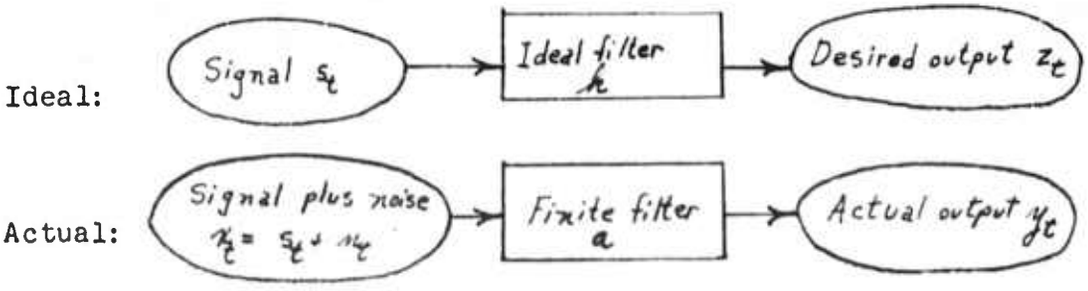
7.2 Solution

With this value of z, the desired spike filter is

$$\check{\beta} = \left[S'S + E\{N'N\}\right]^{-1} S'z$$

8. Time-varying filters

Consider the diagram:



We let T be the present time. We assume:

(1) The signal s_t is of the form

$$s_t = v_t + \sum_{k=1}^{m} c_k f_{kt}$$

where v_t is a random process with $E\{v_t\} = 0$ and known autocovariance $\phi_{vv}(t,s) = E\{v_t v_s\}$, where

$$c_1, c_2, \ldots c_m$$

are unknown but fixed constants, and

$$f_{1t}, f_{2t}, \ldots, f_{mt}$$

are known, non-random functions of time t.

(2) The desired output at time T is

$$z_T = \sum_{t=-\infty}^{\infty} k_t s_t$$

where the coefficients k_t $(-\infty < t < \infty)$ of the ideal filter are known. For example, in case of prediction α units ahead, $k_t = \delta_{T+\alpha,\, t} = \begin{cases} 0 & t \neq T+\alpha \\ 1 & t = T+\alpha \end{cases}$

so $$z_T = \sum_{t=-\infty}^{\infty} \delta_{T+\alpha,t}\, s_t = s_{T+\alpha}$$

$$= \nu_{T+\alpha} + \sum_{k=1}^{m} c_k f_{k,T+\alpha}$$

(3) The actual output is given by (at time T)

$$y_T = \sum_{t=1}^{T} a_t x_t \quad , \quad x_t = s_t + \nu_t .$$

where the coefficients of the finite filter a_1, a_2, ... a_T are to be determined so that

$$E\{(z_T - y_T)^2\} = \text{minimum}$$

under the constraint that the actual output y_T equals the desired output z_T when $x_t = s_t$ and for all possible constants c_1, c_2, ..., c_m.

This constraint may be written

$$z_T = \sum_{t=1}^{T} a_t s_t$$

or

$$\sum_{t=-\infty}^{\infty} k_t s_t = \sum_{t=1}^{T} a_t s_t$$

Letting $c_1 = 1$, $c_2 = 0$, ... $c_m = 0$ we obtain $s_t = f_{1t}$, and so

$$\sum_{t=-\infty}^{\infty} k_t f_{1t} = \sum_{t=1}^{T} a_t f_{1t} .$$

Similarly, we obtain

$$\sum_{t=-\infty}^{\infty} k_t f_{2t} = \sum_{t=1}^{T} a_t f_{2t} \qquad \text{etc.}$$

Hence the constraint is

$$\sum_{t=-\infty}^{\infty} k_t f_{jt} = \sum_{t=1}^{T} a_t f_{jt} \qquad \text{for } j = 1, 2, \ldots, m$$

(4) The actual input is

$$x_t = s_t + n_t$$

where the noise n_t is a random process with zero mean $E\{n_t\} = 0$ and known autocovariance

$$\phi_{nn}(t,s) = E\{n_t n_s\}$$

and known cross-covariance with v_t:

$$\phi_{nv}(t,s) = E\{n_t v_t\}.$$

End of Assumptions

Let us now transform to matrix notation. Write:

1 x T row vectors

$s = (s_1, s_2, \ldots, s_T)$
$n = (n_1, n_2, \ldots, n_T)$
$x = (x_1, x_2, \ldots, x_T)$
$f_j = (f_{j1}, f_{j2}, \ldots, f_{jT})$
$a = (a_1, a_2, \ldots, a_T)$
$v = (v_1, v_2, \ldots, v_T)$

$f = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}$: a mxT matrix

$c = (c_1, c_2, \ldots, c_m)$

1 x ∞ row vectors

$V = (\ldots, v_{-1}, v_0, v_1, \ldots)$
$\mathcal{S} = (\ldots, s_{-1}, s_0, s_1, \ldots)$
$k = (\ldots, k_{-1}, k_0, k_1, \ldots)$
$F_j = (\ldots, f_{j,-1}, f_{j0}, f_{j1}, \ldots)$

$F = \begin{pmatrix} F_1 \\ \vdots \\ F_m \end{pmatrix}$: an m x ∞ matrix

Then: $s = v + cf$

$S = V + cF$ $\qquad z_T = Sk' \qquad y_T = xa'$

$Fk' = fa'$ (the constraint)

$$E\left\{(z_T - y_T)^2\right\} = E\left\{(Sk' - xa')^2\right\} \qquad E(s) = cf \qquad E(S) = cF$$

$x = s + n \qquad E\{x\} = E\{s\} = cf$

$y_T = xa' = (s+n)a' \qquad E\{y_T\} = E\{s\}\,a' = cfa'$ ← =because of constraint

$z_T = Sk' \qquad E\{z_t\} = E\{S\}\,k' = cFk'$ ←

Let us now determine the optimum operator $a = (a_1, \ldots, a_T)$. The mean-square-error is

$$E\{(z_T - y_T)^2\} = E\{(z_T - xa')^2\} \qquad = a'x'xa'$$
$$= E\{z_T^2\} - 2E\{z_T xa'\} + E\{(xa')^2\}$$
$$\equiv \phi_{zz} - 2\phi_{zx}a' + a\phi_{xx}a'$$

where

$$\phi_{zz} = E\{z_T^2\}$$
$$\phi_{zx} = E\{z_T x\} = E\{z_T(x_1, x_2, \ldots, x_T)\}$$
$$= (Ez_T x_1,\ Ez_T x_2, \cdots, Ez_T x_T) \quad : 1 \times T \text{ row vector}$$
$$\phi_{xx} = E(x'x) \quad : T \times T \text{ matrix}$$

The constraint is $Fk' = fa'$. Let $\lambda = (\lambda_1, \ldots, \lambda_m)$ be Lagrangian multipliers. Hence we minimize (with respect to a):

$$J = \phi_{zz} - 2\phi_{zx}a' + a\phi_{xx}a' - 2\lambda[fa' - Fk'] .$$

Thus

$$\frac{\partial J}{\partial a} = -2\phi_{zx} + 2a\phi_{xx} - 2\lambda f = 0 ,$$

or

(1) $$\boxed{\lambda f + \phi_{zx} = a\phi_{xx}}$$

Necessary and sufficient condition for the constrained minimum.

Now: Let $u = v + n$, so that $x = u + cf$, $E(u) = 0$.
Hence

$$\phi_{xx} = E(x'x) = E\{(u' + f'c')(u + cf)\} = E\{u'u\} + f'c'cf$$

or

(2) $$\boxed{\phi_{xx} = \phi_{uu} + f'c'cf \quad \text{where } \phi_{uu} = \phi_{ss} + \phi_{sn} + \phi_{ns} + \phi_{nn}}$$

Now $$\phi_{zx} = E(z'x) = E([Sk]'x) = E\{[(V + cF)k]'x\}$$

$$= E\{[(V + cF)k]'[u + cf]\}$$

$$= E\{k(V' + F'c')(u + cf)\} \overset{\text{since } E(u)=0,\ E(V)=0}{=} kE(V'u) + kF'c'cf$$

or

(3) $$\boxed{\phi_{zx} = k[E\{V'u\} + F'c'cf]}$$

Eq. (1) is $\lambda f + \phi_{zx} = a\phi_{xx}$

The constrains is $Fk' = fa'$ or $\boxed{kF' = af'}$

Subst (2) and (3) gives

$$\lambda f + k\,E(V'u) + k\,F'c'cf = a\,\phi_{uu} + a\,f'c'cf \quad , \text{ or}$$

$$\lambda f + k\,E(V'u) = a\phi_{uu} + \underbrace{(af' - kF')}_{=0 \text{ by the constraint.}} c'cf$$

Thus

$$(4) \qquad \boxed{\lambda f + k\,E(V'u) = a\phi_{uu}}$$

Now <u>define</u> p and q to be

$$f = p\,\phi_{uu} \qquad (p : m \times T \text{ matrix}), \quad p = f\phi_{uu}^{-1}$$

$$k\,E\{V'u\} = q\,\phi_{uu} \qquad (q : 1 \times T \text{ row vector}), \quad q = k\,E\{V'u\}\phi_{uu}^{-1}$$

Then (4) becomes $\lambda p\,\phi_{uu} + q\,\phi_{uu} = a\,\phi_{uu}$

or

$$(5) \qquad \boxed{\lambda p + q = a}$$

This is the desired solution except that we must determine λ. To do so we substitute $a = q + \lambda p$ into the constraint $kF' = af'$. We have

$$kF' = (q + \lambda p)f'$$

which is

$$\lambda pf' = kF' - qf'$$

or

$$\lambda = (kF' - qf')(pf')^{-1}$$

Hence the desired filter is

$$a = q + (kF' - qf')(pf')^{-1}p \quad ,$$

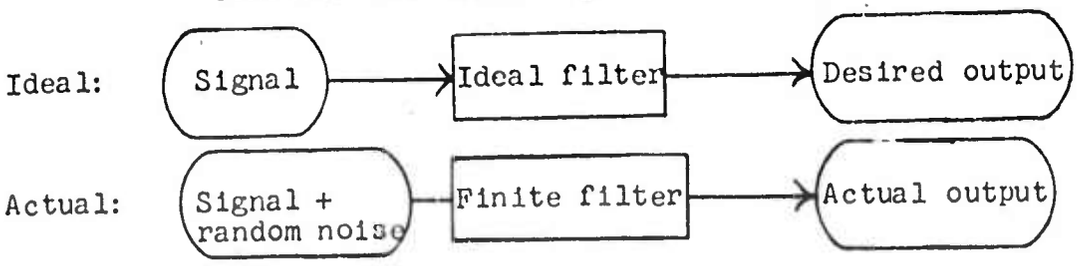
or

$$a = h E\{V'u\} \phi_{uu}^{-1} + (hF' - hE\{V'u\}\phi_{uu}^{-1} f')(f \phi_{uu}^{-1} f')^{-1} f \phi_{uu}^{-1}$$

By letting T vary, we thus obtain the optimum time-varying filter a as a function of T.

9. Detection Filter

Consider the following case:



(1) The Ideal System:

We assume that the signal has the form

$$S_t = \sum_{i=1}^{p} c_i f_{it}$$

where

$$f_{1t}, f_{2t}, \cdots, f_{pt}$$

are p known functions of time t (t an integer) and where

$$c_1, c_2, \cdots, c_p$$

are p unknown constants. The ideal filter is an infinite, non-realizable, time-varying filter described by known coefficients $K_{\tau t}$. The desired output is

$$z_t = \sum_{\tau=-\infty}^{\infty} K_{\tau t} S_t = \sum_{\tau=-\infty}^{\infty} K_{\tau t} \sum_{i=1}^{p} c_i f_{i\tau}$$

$$= \sum_{i=1}^{p} c_i \sum_{\tau=-\infty}^{\infty} K_{\tau t} f_{i\tau} = \sum_{i=1}^{p} c_i g_{it}$$

where $g_{it} \equiv \sum_{\tau=-\infty}^{\infty} f_{i\tau} K_{\tau t}$ is known.

We are interested in the desired output z_t over the time interval t=1,2,...,N, and so we let the 1xN row vector z denote

$$z = (z_1, z_2, \cdots, z_N)$$

Also we let g be the known pxN matrix

$$g = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1N} \\ g_{21} & g_{22} & \cdots & g_{2N} \\ \cdots & & & \cdots \\ g_{p1} & g_{p2} & \cdots & g_{pN} \end{pmatrix}$$

and let c be the unknown 1xp row vector

$$c = (c_1, c_2, \cdots, c_p).$$

Then, clearly, the desired output over the time interval of interest is

$$\boxed{z = cg}$$

where c is unknown and g is known.

(2) The Actual System

We assume that the actual input is

$$x_t = s_t + n_t$$

where s_t is the signal, as given above, and n_t is random noise with zero mean, i.e.

$$E\{n_t\} = 0$$

where $E\{\cdots\}$ denotes ensemble average.

We assume we know the actual input x_t over the time interval t=-M+1,-M+2,...,-1,0,1,2,...,N, and so we let the 1x(M+N) row vector x be

$$x = (x_{-M+1}, x_{-M+2}, \cdots, x_{-1}, x_0, x_1, x_2, \cdots, x_N)$$

We are interested in the desired output z_t over the time interval t=1,2,...,N, and so it follows that we are interested in the actual output y_t over the same time interval. Hence we let the 1xN row vector y be

$$y = (y_1, y_2, \cdots, y_N)$$

Let the Nx(M+N) matrix

$$a = \begin{bmatrix} a_{1,-M+1} & a_{1,-M+2} & \cdots & a_{11} & a_{12} & \cdots & a_{1N} \\ a_{2,-M+1} & a_{2,-M+2} & \cdots & a_{21} & a_{22} & \cdots & a_{2N} \\ & \cdots & & \cdots & & \cdots & \\ a_{N,-M+1} & a_{N,-M+2} & \cdots & a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}$$

denote the finite, time-varying filter, where a is such that input x and output y are related by

$$y = x a'$$

(1×N) (1×(M+N)) ((M+N)×N)

Note Prime indicates matrix transpose.

Since $x_t = s_t + n_t$, we write

$$x = s + n$$

where

$$s = (s_{-M+1}, s_{-M+2}, \cdots, s_{-1}, s_0, s_1, s_2, \cdots, s_N)$$

$$n = (n_{-M+1}, n_{-M+2}, \cdots, n_1, n_0, n_1, n_2, \cdots, n_N)$$

We recall that

$$s_t = \sum_{i=1}^{p} c_i f_{it} \qquad \text{and} \qquad c = (c_1, c_2, \cdots, c_p).$$

Hence define the $p \times (M+N)$ matrix f to be

$$f = \begin{bmatrix} f_{1,-M+1} & f_{1,-M+2} & \cdots & f_{11} & f_{12} & \cdots & f_{1N} \\ f_{2,-M+1} & f_{2,-M+2} & & f_{21} & f_{22} & & f_{2N} \\ & & & & & & \\ f_{p,-M+1} & f_{p,-M+2} & & f_{p1} & f_{p2} & & f_{pN} \end{bmatrix}$$

or

$$f = [f_{it}] \qquad \text{where } i = 1, 2, \cdots, p \text{ and } t = -M+1, -M+2, \cdots, N.$$

Then

$$\underset{1\times(M+N)}{s} = \underset{(1\times p)}{c}\ \underset{p\times(M+N)}{f}$$

The actual output y is therefore

$$y = x a' = (s+n) a' = sa' + na' = cfa' + na',$$

which has ensemble average

$$E\{y\} = E\{cfa'\} + E\{na'\} = cfa'$$

since cfa' is a constant, and

$$E\{n\} = 0.$$

In summary:

The actual output is y = cfa' + na'.
The desired output is z = cg.
The ensemble average of the actual output is $E\{y\} = cfa'$.

Definition The actual output y is said to be unbiased provided that its ensemble average $E\{y\}$ is equal to the desired output z regardless of the value of the weighting factor c, i.e. provided

$$cfa' = cg \qquad \text{regardless of } c.$$

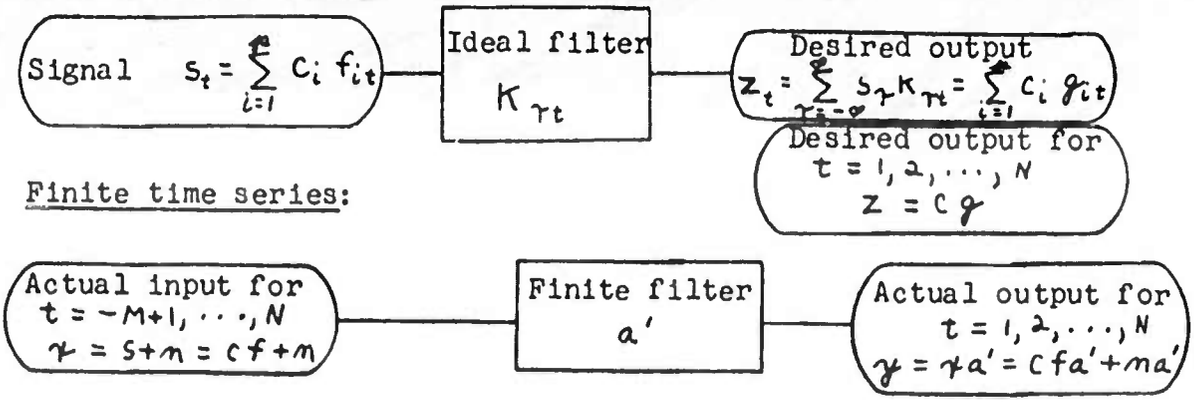
It therefore follows that y is unbiased if and only if

$$fa' = g$$

This is called the unbiasedness constraint

Schematically, we have:

Infinite time series:



Finite time series:

Unbiased condition:

$$E\{\tau\} = z \text{ for all } C \overset{\text{if and only if}}{\Longleftrightarrow} fa' = g$$

(3) Unbiased estimation of the weighting factor c in the case of white noise

Let us first consider the special case in which the random noise n_t is white, i.e.

$$E\{n_t n_s\} = \begin{cases} \sigma^2 & \text{if } t = s \\ 0 & \text{otherwise} \end{cases}$$

where σ^2 is the common variance. Since $x_t = n_t + s_t$ with $E\{n_t\} = 0$, we have

$$\tau = n + s \quad \text{with} \quad E\{\tau\} = s = cf$$

where

$$C = (c_1, \cdots, c_p) \text{ and } f = [f_{it}],\ i = 1, \cdots, p \text{ and } t = -M+1, \cdots, N.$$

Thus we have M+N observations

$$\tau = (\tau_{-M+1}, \cdots, \tau_N)$$

concerning which we make the following assumptions:

(1) Their ensemble averages are linear combinations of the p unknown parameters $c=(c_1, \ldots, c_p)$ Thus

$$E\{x_t\} = \sum_{i=1}^{p} c_i f_{it} \qquad (t = -M+1, \ldots, N)$$

or

$$E\{x\} = cf$$

where the matrix f of order $p \times (M+N)$ is <u>known</u> and is termed the <u>design matrix</u>, and where the row vector c is <u>unknown</u>.

(2) Their covariance matrix is the product of an <u>unknown scalar</u> σ^2 and a known positive definite matrix $\phi = [\phi_{tr}]$. Thus

$$cov\{x_t, x_r\} = \sigma^2 \phi_{tr} = E\{(x_t - E\{x_t\})(x_r - E\{x_r\})\}$$

$$= E\{(x_t - s_t)(x_r - s_r)\} = E\{n_t n_r\}$$

or

$$\sigma^2 \phi = E\{(x-s)'(x-s)\} = E\{n'n\}$$

(3) In the special case now under discussion, $\phi = I$ = identity matrix, i.e. each observation x_t has the same variance σ^2 and every pair x_t, x_r is uncorrelated $(t \neq r)$.

(4) We assume the rank of f is p, and that $p < M+N$.

According to the principle of least-squares, we estimate $c_1, c_2, \cdots, c_p$ simultaneously by selecting those functions $\hat{c}_1, \hat{c}_2, \cdots \hat{c}_p$ of $x_{-M+1}, \cdots, x_1, x_2, \cdots x_N$ which minimize

$$J = \sum_{t=-M+1}^{N} \left(x_t - \sum_{i=1}^{p} c_i f_{it} \right)^2 = (x - cf)(x-cf)'$$

with respect to $c_1, c_2, \ldots, c_p$ considered as independent variables. We have, by differentiating J with respect to c and equating to zero,

$$\frac{\partial J}{\partial c} = 2(x - \hat{c}f)(-f') = 0 \quad , \quad \text{or}$$

$$\boxed{\hat{c}\,ff' = xf'}$$

which are the normal equations, or equations of estimation for the parameter c. Since f has rank p, there is no non-null vector v such that $vf = 0$ and therefore $(vf)(vf)' = vff'v$ is positive, so ff' is positive definite. Hence the normal equations have a unique solution.

Since

$$\frac{\partial J}{\partial c} = -2(xf' - c\,ff')$$

we have

$$\frac{\partial^2 J}{\partial c^2} = 2ff'$$

That is, the matrix of second-order differential coefficients of J with respect to c is 2ff', which is positive definite, so J has an absolute minimum when $c = \hat{c} = (xf')(ff')^{-1}$.

Now let us consider the ensemble average of $\hat{c}$. We have

$$E\{\hat{c}\} = E\{(xf')(ff')^{-1}\} = E\{x\}\, f'(ff')^{-1}$$

But $E\{x\} = s = cf$. Hence

$$E\{\hat{c}\} = c\,ff'(ff')^{-1} = c .$$

Hence the estimate c is <u>unbiased</u>, in the sense that

$$E\{\hat{c}\} = c.$$

The covariance matrix of $\hat{c}$ is

$$\begin{aligned}
\operatorname{cov}(\hat{c},\hat{c}) &= \operatorname{cov}\left(x f'(ff')^{-1}, x f'(ff')^{-1}\right) \\
&= E\left\{\left[(x-s)\, f'(ff')^{-1}\right]'\left[(x-s) f'(ff')^{-1}\right]\right\} \\
&= (ff')^{-1} f\, E\{n'n\}\, f'(ff')^{-1} \\
&= \sigma^2 (ff')^{-1} ff'(ff')^{-1} \\
&= \sigma^2 (ff')^{-1}.
\end{aligned}$$

The least-squares estimate $\hat{c}$ of c is a linear function of the observations x, i.e.

$$\hat{c} = x\left[f'(ff')^{-1}\right] \equiv x\lambda \quad \text{where we define } \lambda \equiv f'(ff')^{-1}$$

Consider any other linear estimate of c, say

$$\tilde{c} = xL \qquad (L \text{ is any matrix of order } (M+N) \times p)$$

Because

$$E\{\tilde{c}\} = E\{xL\} = E\{x\}L = cfL,$$

we see that $\tilde{c}$ is unbiased for all c provided that

$$fL = I$$

The covariance matrix of $\tilde{c}$ is

$$\begin{aligned}
\operatorname{cov}(\tilde{c},\tilde{c}) &= \operatorname{cov}\{xL, xL\} = L' \operatorname{cov}\{x,x\} L \\
&= \sigma^2 L'L.
\end{aligned}$$

We now wish to show that the least-squares estimate $\hat{c}$ is better than any other unbiased linear estimate $\tilde{c}$ in the sense that, for each parameter c_j,

$$\text{var } \hat{c}_j \leq \text{var } \tilde{c}_j$$

Proof We have

$$\lambda'\lambda + (L-\lambda)'(L-\lambda) = 2\lambda'\lambda - L'\lambda - \lambda'L + L'L$$

We recall that

$$\lambda = f'(ff')^{-1} \text{ and that } fL = I.$$

Hence

$$L'\lambda = L'f'(ff')^{-1} = I(ff')^{-1} = (ff')^{-1} \qquad (1)$$

Also

$$\lambda'L = (L'\lambda)' = [(ff')^{-1}]' = [(ff')']^{-1} = (ff')^{-1} \qquad (2)$$

Also

$$\lambda'\lambda = (ff')^{-1} f f'(ff')^{-1} = (ff')^{-1} \qquad (3)$$

Using (1), (2), (3), we have

$$\lambda'\lambda + (L-\lambda)'(L-\lambda) = 2(ff')^{-1} - (ff')^{-1} - (ff')^{-1} + L'L$$
$$= L'L.$$

Hence

$$\underbrace{\sigma^2 L'L} = \underbrace{\sigma^2\lambda'\lambda} + \sigma^2(L-\lambda)'(L-\lambda)$$

$$\text{cov}(\tilde{c},\tilde{c}) = \text{cov}(\hat{c},\hat{c}) + \sigma^2(L-\lambda)'(L-\lambda)$$

Each diagonal element of $\text{cov}(\tilde{c},\tilde{c})$ is therefore minimized if the corresponding column of $L-\lambda$ consists entirely of zeros.

Hence $\hat{c}$ is the best linear unbiased estimate of c.

QED

(4) Unbiased estimation of linear combinations of the weighting factor c in the case of white noise

We now wish to consider unbiased linear estimates of the desired output z=cg, where c is the unknown weighting factor and g is a known pXN matrix (see page). Suppose that $\tilde{z} = x\ell$ is an unbiased estimate of z. That is,

$$cg = E\{\tilde{z}\} = E\{x\ell\} = E\{x\}\ell = cf\ell$$

whatever c. Then

$$g = f\ell .$$

The covariance matrix of z is

$$cov(\tilde{z},\tilde{z}) = cov(x\ell, x\ell) = \ell' cov(x,x)\ell = \sigma^2 \ell'\ell .$$

Consider now the estimate $\hat{z} = \hat{c}g = x\lambda g$ where $\hat{c}$ is the best unbiased linear estimate of c. Its covariance matrix is

$$cov(\hat{z},\hat{z}) = cov(\hat{c}g, \hat{c}g) = g' cov(\hat{c},\hat{c}) g$$
$$= g' \sigma^2 (ff')^{-1} g = \sigma^2 g'\lambda'\lambda g$$

Now

$$(\lambda g)'(\lambda g) + (\ell - \lambda g)'(\ell - \lambda g) = 2(\lambda g)'(\lambda g) - \ell'\lambda g - g'\lambda'\ell + \ell'\ell$$
$$= 2g'(ff')^{-1} g - \underbrace{\ell' f'}_{g'}(ff')^{-1} g - g'(ff')^{-1} \underbrace{f\ell}_{g} + \ell'\ell$$
$$= \ell'\ell .$$

Therefore

$$\sigma^2 \ell'\ell = \sigma^2 (\lambda g)'(\lambda g) + \sigma^2 (\ell - \lambda g)'(\ell - \lambda g)$$

or

$$\mathrm{cov}(\tilde{z},\tilde{z}) = \mathrm{cov}(\hat{z},\hat{z}) + \sigma^2 (\ell-\lambda g)'(\ell-\lambda g)$$

Each diagonal element of $\mathrm{cov}(\hat{z},\hat{z})$ is always less than or equal to the corresponding diagonal element of $\mathrm{cov}(\tilde{z},\tilde{z})$, which shows that

$$\hat{z} = \hat{c} g = x \lambda g$$

where $$\hat{c} = x\lambda = x f' (ff')^{-1}$$

is the best unbiased linear estimate of the desired output

$$z = cg.$$

(5) Unbiased linear estimation of the weighting factor c in the case of colored noise.

We now drop the assumption that the noise is white, that is, we drop assumption (3) on page

Let T denote the lower triangle matrix such that $\phi = T'T$ and define $u = xT^{-1}$.

Then

$$Eu = E\{xT^{-1}\} = E\{x\}T^{-1} = cfT^{-1}$$

and

$$\begin{aligned}\mathrm{cov}(u,u) &= (T^{-1})' \mathrm{cov}(x,x)\, T^{-1} = (T^{-1})' \sigma^2 \phi\, T^{-1} \\ &= (T')^{-1} \sigma^2 T' T T^{-1} = \sigma^2 (T^{-1})' T' T T^{-1} \\ &= \sigma^2 (TT^{-1})' T T^{-1} = \sigma^2 I,\end{aligned}$$

so $u = xT^{-1}$ is white noise.

Since T^{-1} is non-singular, fT^{-1} has the same rank as f, which is p. Using the results on page , we see that the best unbiased linear estimate of c is

$$\hat{c} = \mathcal{X}\left[(fT^{-1})'\{(fT^{-1})(fT^{-1})'\}^{-1}\right]$$

$$= \mathcal{X}T^{-1}\left[(T^{-1})'f'\{fT^{-1}T^{-1\prime}f'\}^{-1}\right]$$

$$= \mathcal{X}(T^{-1}T^{-1\prime})\, f'\{f(T^{-1}T^{-1\prime})f'\}^{-1}$$

$$= \mathcal{X}(T'T)^{-1}\, f'\{f(T'T)^{-1}f'\}^{-1}, \text{ or}$$

$$\boxed{\hat{c} = \mathcal{X}\phi^{-1}f'\{f\phi^{-1}f'\}^{-1}}$$

The covariance matrix of $\hat{c}$ is

$$\text{cov}(\hat{c},\hat{c}) = \sigma^2\left[fT^{-1}(fT^{-1})'\right]^{-1}$$

$$= \sigma^2\left[fT^{-1}T^{-1\prime}f'\right]^{-1}, \text{ or}$$

$$\boxed{\text{cov}(\hat{c},\hat{c}) = \sigma^2(f\phi^{-1}f')^{-1}}$$

The estimate $\hat{c}$ is the value of c at which the quadratic form

$$(\mathcal{X}-cf)\,\phi^{-1}(\mathcal{X}-cf)'$$

attains its minimum.

(6) Unbiased linear estimation of linear combinations of the weighting factor c in the case of colored noise

The best unbiased linear estimate of $z = cg$ is

$$\hat{z} = \hat{c}\, g = \kappa \phi^{-1} f' \{ f \phi^{-1} f' \}^{-1} g$$

with covariance matrix

$$\mathrm{cov}(\hat{z}, \hat{z}) = \sigma^2 g' \{ f T^{-1} T^{-1\prime} f' \}^{-1} g \quad , \text{ or}$$

$$\mathrm{cov}(\hat{z}, \hat{z}) = \sigma^2 g' [f \phi^{-1} f']^{-1} g$$

(7) Determination of the finite filter a' subject to the unbiased condition

We wish to find a' such that E(z-y)'(z-y) is a minimum subject to $fa' = g$. We have

$$J = E\{(z-y)'(z-y)\} = E\{(cg - cfa' - \kappa a')'(cg - cfa' - \kappa a')\}$$
$$= E\{(\kappa a')'(\kappa a')\} = a\, E\{\kappa' \kappa\} a' = \sigma^2 a \phi a'$$

Introduce the Lagrange multipliers $\lambda = (\lambda_1, \cdots, \lambda_p)$ and minimize $J + 2\lambda(g - fa')$. Setting the derivative of this equal to zero we obtain $\sigma^2 \phi = \lambda f$. Solving the simultaneous equations for a and λ :

$$\begin{cases} a \sigma^2 \phi = \lambda f \\ f a' = g \end{cases}$$

we obtain

$$a = g' (f \phi^{-1} f')^{-1} f \phi^{-1} \qquad \lambda = g' (f \sigma^2 \phi^{-1} f')^{-1}$$

Hence the actual output is

$$y = \kappa a' = \kappa \phi^{-1} f' (f \phi^{-1} f')^{-1} g = \hat{z} = \hat{c}\, g$$ as given in subsection (6) above.

10. Computational Aides

Suppose we wish to divide the z-transform

$$\Pi(z) = a_0 z^{m} + a_1 z^{m-1} + a_2 z^{m-2} + \cdots + a_r z^{m-r} + \cdots + a_{m-1} z + a_m$$

of the finite operator $a_0, a_1, \cdots, a_m$ by the z-transform

$$F(z) = \alpha_0 z^{n} + \alpha_1 z^{n-1} + \alpha_2 z^{n-2} + \cdots + \alpha_{n-1} z + \alpha_n$$

of the finite operator $\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_n$. We have

$$\frac{\Pi(z)}{F(z)} = A_0 z^{m-n} + A_1 z^{m-n-1} + A_2 z^{m-n-2} + \cdots + A_r z^{m-n-r} + \cdots .$$

Then

$$A_0 = \frac{1}{\alpha_0} \left| a_0 \right| ,$$

$$A_1 = \frac{1}{\alpha_0^2} \begin{vmatrix} \alpha_0 & a_0 \\ \alpha_1 & a_1 \end{vmatrix} ,$$

$$A_2 = \frac{1}{\alpha_0^3} \begin{vmatrix} \alpha_0 & 0 & a_0 \\ \alpha_1 & \alpha_0 & a_1 \\ \alpha_2 & \alpha_1 & a_2 \end{vmatrix} ,$$

$$A_3 = \frac{1}{\alpha_0^4} \begin{vmatrix} \alpha_0 & 0 & 0 & a_0 \\ \alpha_1 & \alpha_0 & 0 & a_1 \\ \alpha_2 & \alpha_1 & \alpha_0 & a_2 \\ \alpha_3 & \alpha_2 & \alpha_1 & a_3 \end{vmatrix} ,$$

$$A_4 = \frac{1}{\alpha_0^5} \begin{vmatrix} \alpha_0 & 0 & 0 & 0 & a_0 \\ \alpha_1 & \alpha_0 & 0 & 0 & a_1 \\ \alpha_2 & \alpha_1 & \alpha_0 & 0 & a_2 \\ \alpha_3 & \alpha_2 & \alpha_1 & \alpha_0 & a_3 \\ \alpha_4 & \alpha_3 & \alpha_2 & \alpha_1 & a_4 \end{vmatrix} ,$$

and in general

$$A_n = \frac{1}{\alpha_0 \alpha_{n+1}} \begin{vmatrix} \alpha_0 & 0 & 0 & 0 & \cdots & a_0 \\ \alpha_1 & \alpha_0 & 0 & 0 & & a_1 \\ \alpha_2 & \alpha_1 & \alpha_0 & 0 & & a_2 \\ \alpha_3 & \alpha_2 & \alpha_1 & \alpha_0 & \cdots & a_3 \\ \vdots & & & \cdots & & \\ \alpha_n & \alpha_{n-1} & \alpha_{n-2} & \alpha_{n-3} & \cdots & a_n \end{vmatrix}$$

$|\alpha_0| \neq 0$ must be small

where $\left| \quad \right|$ indicates determinent.

Example 1: Divide $5z^6 + 3z^5 + z^4 + 2z^3 + 3z^2 + 6z + 7$ by $2z^4 + 3z^3 + 4z^2 + 5z + 7$. These coefficients are (5, 3, 1, 2, 3, 6, 7) and (2, 3, 4, 5, 7). The quotient is

$$A_0 z^2 + A_1 z + A_2$$

where

$$A_0 = \frac{5}{2} \quad , A_1 = \frac{1}{4}\begin{vmatrix} 2 & 5 \\ 3 & 3 \end{vmatrix} \quad , A_2 = \frac{1}{8}\begin{vmatrix} 2 & 0 & 5 \\ 3 & 2 & 3 \\ 4 & 3 & 1 \end{vmatrix}$$

or $A_0 = \frac{5}{2}$, $A_1 = -\frac{9}{4}$, $A_2 = -\frac{9}{8}$.

11. Squared-magnitude devices

11.1 Assumptions

(a) The finite time-series

$$x_0, x_1, \ldots, x_N$$

represents (N+1) consecutive observations from a regular stationary stochastic process with zero mean and with autocovariance

$$\phi(r) = E\{x_{t+r} \cdot x_t\}$$

and spectral density $\Phi(\omega)$.

(b) The infinitely long time-series

$$\ldots, 0, 0, x_0, x_1, \ldots, x_N, 0, \ldots$$

formed by letting $x_t = 0$ for $t < 0$ and for $t < N$ is the input to system with impulse response b_t. The impulse response b_t may be infinitely long in both directions, i.e., b_t may be different from zero for both $t \to \infty$ and $t \to -\infty$.

(c) The output time-series y_t, which in general will be infinitely long in both time directions, is

$$y_t = \sum_{\lambda=-\infty}^{\infty} b_\lambda x_{t-\lambda} \qquad (-\infty < t < \infty)$$

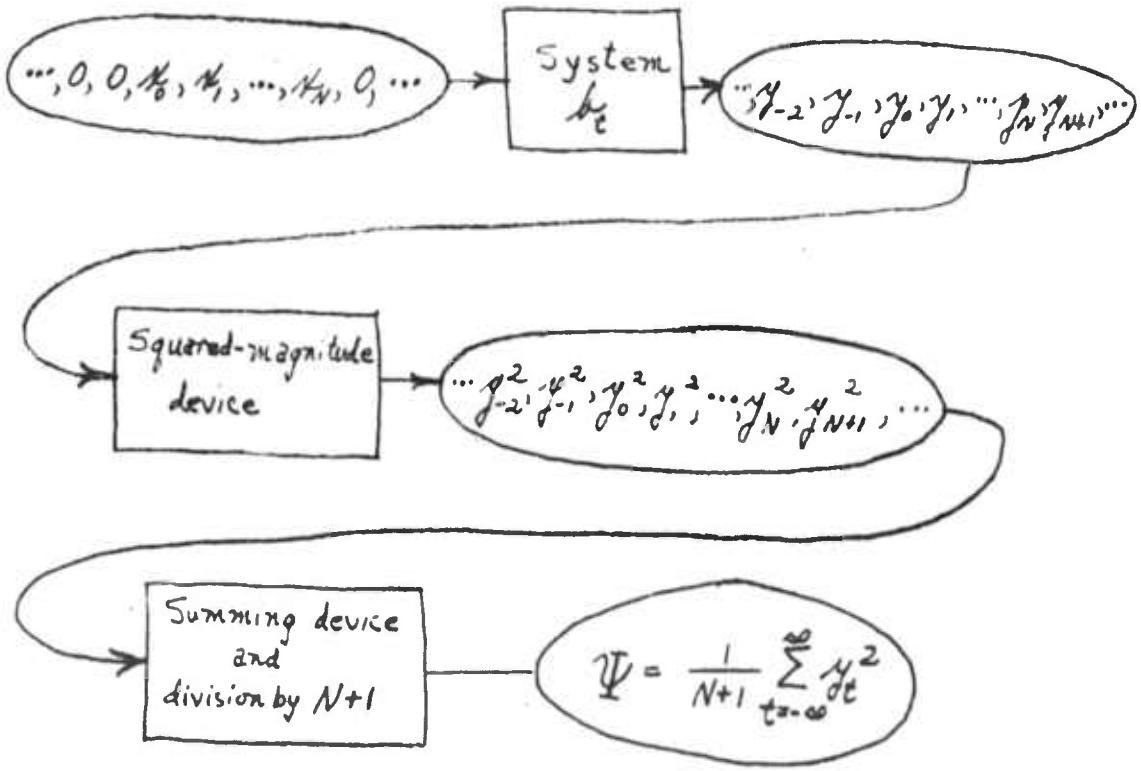
(d) The output y_t is the input to a squared-magnitude device whose output is y_t^2.

(e) Finally, y_t^2 is the input to a device that sums y_t^2 from $t = -\infty$ to $t = \infty$ and then

divides the result by N+1 . Thus the output, denoted by Ψ , is

$$\Psi = \frac{1}{N+1} \sum_{t=-\infty}^{\infty} y_t^2 .$$

Schematically, we have



12. References

Simpson, S.M. (1961-1963) Initial Studies on Underground Nuclear Detection with Seismic Data Prepared by a Novel Digitization System, Report. No. 1
Time Series Techniques Applied to Underground Nuclear Detection and Further Digitized Seismic Data, Report No.2
Continued Numerical Studies on Underground Nuclear Detection and Further Digitized Seismic Data Report No. 3
Magnetic Tape Copies of MIT Geophysics Program Set I, (Time Series Programs for the IBM 709, 7090), Report No. 4
Air Force Cambridge Research Center, Project Vela Uniform

Robinson, E.A., (1962), Random Wavelets and Cybernetic Systems, Stechert-Hafner, New York, and Griffin, London

3. Normal Equation Forms in Optimum Filtering Problems

In section 2 of this report the formal equations for several types of filters (matched, time-invariant least square, time varying, etc.) were developed. In sections 3 and 4, we will concentrate on the time-invariant least square filter. However, we will generalize this case to include both multi-input, and multi-dimensional processes.

Section 3 develops the normal equations for single, multi-input, and multi-dimensional process optimum filters. Highly efficient computational schemes for solving these equations are given in section 4.

The transient autocorrelation of a complex single process is standardly written as

$$r_t = \sum_{i=1}^{N} X_{i+t}\, \bar{X}_i \qquad (3.-1)$$

where $\bar{X}$ indicates complex conjugate.
If we think of X as a matrix valued process (real or complex) and $\bar{X}$ as the transpose of X, then the autocorrelation of X can have the same formal definition. This convention will be used throughout section 3. and 4. in order to preserve the formal similarity of the different systems. The reader will note that this alters the form of the single process development as was given in section 2.

3.1 Single Process

The single process is characterized by <u>one</u> set of numbers corresponding to discrete intervals of time. Thus, the series

$$s_t = 1., .5, .25, .125, \ldots \qquad t = 1, 2, \ldots$$

corresponds to the signal

$$s(t) = \begin{cases} 0 & t < 1 \\ 1/t & t \geq 1 \end{cases}$$

In this section we will determine the optimum least-square filter for operating on such a process.

3.1.1 Assumptions

a) The signal s_t has a known fixed shape.

b) The noise n_t is a random process with unknown distribution.

c) The mean value $E\{n_t\}$ of the noise is known to be zero.

d) The auto-covariance of the noise $E\{n_i n_j\}$ is known.

e) Time is a discrete integer valued parameter.

f) The observed random process is

$$x_t = s_t + n_t .$$

g) The random process is observed for $t = 0, 1, 2 \ldots N$.

h) The observed random process in convolved with the coefficients of the impulse response of a linear filter $f_1 \ldots f_M$ $M < N$ (to be determined).

i) The actual output of the filter is

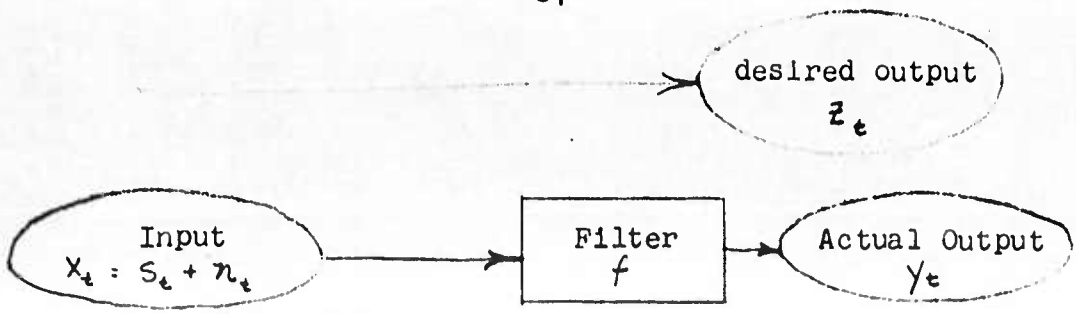
$$y_t = \sum_{s=1}^{M} f_s \, x_{t-s} \qquad t = M-1, M, \ldots, N .$$

j) The desired output of the filter is z_t $t = M-1, M, \ldots N$ where z_t is a known fixed function.

3.1.2 Statement of Problem

Determine f such that the sum of the errors squared is a minimum:

$$E\left\{ \sum_{t=M-1}^{N} (z_t - y_t)^2 \right\} = \text{minimum} \qquad (3.1\text{-}1)$$



3.1.3 Matrix Notation

$$S = \begin{bmatrix} s_{M-1} & s_M & \cdots & s_N \\ s_{M-2} & s_{M-1} & \cdots & s_{N-1} \\ \vdots & \vdots & & \vdots \\ s_0 & s_1 & \cdots & s_{N-M+1} \end{bmatrix} \quad \begin{matrix} \longleftarrow N-M+2 \longrightarrow \\ \\ \\ \\ \uparrow \\ M \\ \downarrow \end{matrix} \tag{3.1-2}$$

$$N = \begin{bmatrix} n_{M-1} & n_M & \cdots & n_N \\ n_{M-2} & n_{M-1} & \cdots & n_{N-1} \\ \vdots & \vdots & & \\ n_0 & n_1 & \cdots & n_{N-M+1} \end{bmatrix} \tag{3.1-3}$$

$$X = S + N \tag{3.1-4}$$

$$Y = \underset{\longleftarrow N-M+2 \longrightarrow}{\begin{bmatrix} y_{M-1} & \cdots & y_N \end{bmatrix}} \tag{3.1-5}$$

$$z = \begin{bmatrix} z_{M-1} & \cdots & z_N \end{bmatrix} \tag{3.1-6}$$

$$e = \begin{bmatrix} e_{M-1} & \cdots & e_N \end{bmatrix} \tag{3.1-7}$$

$$= z - y \tag{3.1-8}$$

$$f = \underset{\longleftarrow M \longrightarrow}{\begin{bmatrix} f_1 & \cdots & f_M \end{bmatrix}} \tag{3.1-9}$$

Thus

$$f X = y \tag{3.1-10}$$

3.1.4 Determination of the filter f.

Let α be the expected value of the sum of the squared errors

$$\alpha \equiv E\{e\bar{e}\} \qquad \bar{e} \Rightarrow e \text{ transpose} \tag{3.1-11}$$

$$= E\left\{(z - fX)\overline{(z - fX)}\right\} \tag{3.1-12}$$

Now, if we take the derivative of α with respect to f and equate it to zero in order to minimize α, we find

$$\frac{\partial \alpha}{\partial f} = 0 \Rightarrow E\{(z - fX)\bar{X}\} = 0 \tag{3.1-13}$$

or

$$E\{e\bar{X}\} = 0 \tag{3.1-14}$$

Thus, the condition that α be minimized is equivalent to saying that the error e must be normal to the process X. For this reason, this equation is known as the normal equation for f.

Now expand the terms in the normal equation:

$$E\{z\bar{X} - fX\bar{X}\} = 0. \tag{3.1-15}$$

But,

$$X = S + N$$

and

$$E\{N\} = 0.$$

Thus

$$E\left\{ z(\overline{S+N}) - f(S+N)(\overline{S+N}) \right\} = 0 \qquad (3.1\text{-}16)$$

$$f\left(S\bar{S} + E\left\{ N\bar{N} \right\} \right) = z\,\bar{S} \qquad (3.1\text{-}17)$$

If we examine the multiplication $S\bar{S}$ in detail

$$S\bar{S} = \begin{bmatrix} s_{M-1} & s_M & \cdots & s_N \\ s_{M-2} & s_{M-1} & \cdots & s_{N-1} \\ \vdots & \vdots & & \vdots \\ s_0 & s_1 & \cdots & s_{N-M+2} \end{bmatrix} \begin{bmatrix} \bar{s}_{N-1} & \bar{s}_{N-2} & \cdots & \bar{s}_0 \\ \bar{s}_M & \bar{s}_{M-1} & \cdots & \bar{s}_1 \\ \vdots & \vdots & & \vdots \\ \bar{s}_N & \bar{s}_{N-1} & \cdots & \bar{s}_{N-M+2} \end{bmatrix} \qquad (3.1\text{-}18)$$

we see that the multiplication of the first row of S by the first column of $\bar{S}$, the second row by the second column, etc., is like the 0 lag of a transient autocorrelation of a portion of s. Since each multiplication is taken over different limits, the terms along a diagonal of $S\bar{S}$ will not be the same.

For ease in computation, we desire that the term $S\bar{S} + E\left\{N\bar{N}\right\}$ be Toeplitz, i.e., that the elements along each diagonal be the same. This can be accomplished in two ways:

1) If $N \gg M$ and s is a stationary random process, then we make the approximation that

$$\frac{1}{N-M+2} \sum_{i=M-1}^{N} s_{i-j}\,\bar{s}_i \simeq E\left\{ s_{i-j}\,\bar{s}_i \right\} \qquad (3.1\text{-}19)$$

The normal equation becomes

$$f\left[E\left\{ S\bar{S} + N\bar{N} \right\} \right] = E\left\{ z\,\bar{S} \right\} \qquad (3.1\text{-}20)$$

or

$$f\,R = g$$

where $$R = E\{S\bar{S} + N\bar{N}\} \tag{3.1-21}$$
$$g = E\{z\bar{S}\}$$

2) We assume s is a transient. We then redefine the matrix S to be

$$S' = \begin{bmatrix} s_0 & s_1 & \cdots & s_N & 0 & \cdots & 0 \\ 0 & s_0 & \cdots & s_{N-1} & s_N & 0 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdots & 0 & s_0 & s_1 & \cdots & s_N \end{bmatrix} \tag{3.1-22}$$

and the outputs

$$y' = [y_0 \; y_1 \; \cdots \; y_{N+M-1}] \tag{3.1-23}$$

$$z' = [z_0 \; z_1 \; \cdots \; z_{N+M-1}] \tag{3.1-24}$$

$$e' = [e_0 \; e_1 \; \cdots \; e_{N+M-1}] \tag{3.1-25}$$

The normal equation now becomes

$$f\left[S'\bar{S}' + E\{N\bar{N}\}\right] = z'\bar{S}' \tag{3.1-26}$$

or $$f\,R = g \tag{3.1-27}$$

where $$R = S'\bar{S}' + E\{N\bar{N}\} \tag{3.1-28}$$

$$g = z'\bar{S}' \tag{3.1-29}$$

$$R = \begin{bmatrix} r_0 & r_1 & \cdots & r_{M-1} \\ r_{-1} & r_0 & \cdots & r_{M-2} \\ \vdots & & & \vdots \\ r_{-M+1} & r_{-M+2} & \cdots & r_0 \end{bmatrix} \tag{3.1-30}$$

$$r_i = \overline{r_{-i}} \tag{3.1-31}$$

$$g = [\, g_0 \; g_1 \cdots g_n \,] \tag{3.1-32}$$

3.1.5 Determination of the expected error α.

From equation (3.1-11) we defined α to be

$$\alpha = E\{ e\, \bar{e} \} \tag{3.1-11a}$$

If we substitute $e = z - fX$, we find

$$\alpha = E\{ e\, \overline{(z - fX)} \} \tag{3.1-33}$$

But, since e is normal to x and f is a linear operator, we get

$$\alpha = E\{ e\, \bar{z} \}$$

$$= E\{ (z - fX)\, \bar{z} \}$$

$$= E\{ z\bar{z} - fX\bar{z} \}$$

$$= z\bar{z} - f\bar{g} \tag{3.1-34}$$

since $E\{n\} = 0$.

3.1.6 Prediction

A special case of interest is that of predicting future values of a series from past values. For this case we set the desired output to be the signal at some future time:

$$z = s_k \tag{3.1-35}$$

$$[\, z_0 \; z_1 \; \ldots \; z_N \,] = [\, s_k \; s_{k+1} \; \ldots \; s_{N+k} \,]$$

where k is the prediction distance. The normal equation becomes

$$f\left[S\bar{S}' + E\{N\bar{N}\} \right] = s_k \bar{S} \tag{3.1-36}$$

Now, if there is no noise, this takes on the form:

$$\left[f_1 \; f_2 \ldots f_M \right] \begin{bmatrix} r_0 & r_1 & \cdots & r_{M-1} \\ r_1 & r_0 & \cdots & r_{M-2} \\ \vdots & & & \vdots \\ r_{M-1} & r_{M-2} & \cdots & r_0 \end{bmatrix} = \left[r_k \; r_{k+1} \ldots r_{k+M} \right] \tag{3.1-37}$$

3.1.7 Prediction Error

In many cases we are more interested in the error involved in predicting rather than the actual prediction. Equation (3.1-8) defined the error to be the difference between the desired output and the actual output. This can be written as

$$\begin{aligned} e &= z - fX \\ &= s_k - fS \end{aligned} \tag{3.1-38}$$

since we have assumed that there is no noise involved. If we define

$$f' = \left[1, \underbrace{0, \ldots, 0}_{k-1 \text{ ZEROS}}, -f_1, -f_2 \ldots -f_M \right] \tag{3.1-39}$$

and expand S suitably, then the error can be written

$$e = f'S \tag{3.1-40}$$

3.2 Multi-Input Processes

The multi-input process is characterized by having several separate time series. These time series are arranged so that each point in time is represented by a column matrix where each term in the matrix corresponds to a specific time series. This leads to the idea of a matrix-valued time series for which we wish to find a matrix-valued least-square optimum filter.

3.2.1 Assumptions

a) The $n \times 1$ matrix-valued signal s_t has a known fixed shape

$$s_t = \begin{bmatrix} s_t^1 \\ \cdot \\ \cdot \\ \cdot \\ s_t^n \end{bmatrix} .$$

b) The $n \times 1$ matrix valued noise n_t is a random process with unknown distribution.

c) The mean value of the noise $E\{n_i\}$ is known to be zero.

d) The covariance of the noise $E\{n_i \bar{n}_j\}$ is known.

e) Time is a discrete integer valued parameter.

f) The observed random process is

$$x_t = s_t + n_t$$

g) The random process is observed for $t = 0, 1, \ldots N$

h) The observed random process is convolved with a $l \times n$ $l \leq n$ matrix-valued linear filter $f_1 \ldots f_m$ $m < N$ (to be determined).

i) The $l \times 1$ $l \leq n$ matrix-valued actual output of the filter is

$$y_t = \sum_{s=1}^{M} f_s x_{t-s} \qquad t = M-1, M, \ldots N$$

j) The $l \times 1$ $l \leq n$ matrix-valued desired output is z_t $t = M-1, M, \ldots, N$ where z_t is a known fixed function.

Using these assumptions we can now determine the filter f by following the development for single processes (3.1.3-3.1.7). Each term of the matrices that are defined there will now be a matrix rather than a scalar. This is essentially only an interior grouping of terms within the matrix. If this grouping is removed, leaving the individual scalar terms arranged as they were, the matrix will have a normal configuration and interpretation.

The normal equation is

$$\begin{bmatrix} f_1 & f_2 & \cdots & f_m \end{bmatrix} \begin{bmatrix} r_0 & r_1 & \cdots & r_{m-1} \\ r_{-1} & r_0 & \cdots & r_{m-2} \\ \vdots & & & \vdots \\ r_{-m+1} & r_{-m+2} & \cdots & r_0 \end{bmatrix} = \begin{bmatrix} g_1 & g_2 & \cdots & g_m \end{bmatrix} \tag{3.2-1}$$

where the expected error is

$$\alpha = z\bar{z} - f\bar{g} \tag{3.2-2}$$

Now r_i is a $n \times n$ matrix which contains all terms of the i^{th} lag of the autocorrelations and crosscorrelations of the input series. (Note that $r_i = \bar{r}_{-i}$). Likewise g_i is a $l \times n$ matrix.

The restriction in the assumption 3.2- h), i), and j) that $l \leq n$ follows from the fact that we can make only n linearly independent combinations of the inputs.

3.3 Multi-Dimensional Processes

A multi-dimensional process will be characterized by a multi-dimensional data array. In two dimensions a x^{i_1,i_2} array might have the form

$$\begin{matrix} \mathcal{x}^{1,1} & \mathcal{x}^{1,2} & \cdots & \mathcal{x}^{1,m_2} \\ \mathcal{x}^{2,1} & \mathcal{x}^{2,2} & \cdots & \mathcal{x}^{2,m_2} \\ \vdots & \vdots & & \\ \mathcal{x}^{m_1,1} & \mathcal{x}^{m_1,2} & \cdots & \mathcal{x}^{m_1,m_2} \end{matrix}$$

(We make use of script symbols in order to emphasize the multi-dimensionality of the process.)

For some applications, one of these directions may be thought of as time. With no loss of generality, we will suppress this interpretation for the development.

3.3.1 Terminology

a) Dot Product

$$\left[\mathcal{x}, \bar{y} \right] = \bar{\sum_{i_1, i_N = -\infty}} \mathcal{x}^{i_1 \cdots i_N} \bar{y}^{i_1 \cdots i_N} \tag{3.3-1}$$

b) Displaced Dot Product

$$\left[\mathcal{x}^{i_1 \cdots i_N}, y^{j_1 + i_1, \cdots i_N + j_N} \right] = \bar{\sum_{i_1 \cdots i_N = -\infty}} \mathcal{x}^{i_1 \cdots i_N} \, y^{i_1 + j_1 \cdots i_N + j_N} \tag{3.3-2}$$

c) Reversed Dot Product

$$\left[\mathcal{x}^{i_1 \cdots i_N}, y^{j_1 - i_1, \cdots j_N - i_N} \right] = \bar{\sum_{i_1 \cdots i_N = -\infty}} \mathcal{x}^{i_1 \cdots i_N} \, y^{j_1 - i_1 \cdots j_N - i_N} \tag{3.3-3}$$

d) Convolution

$$z^{j_1 \cdots j_N} = \left[x^{i_1 \cdots i_N} , y^{j_1 - i_1 \cdots j_N - i_N} \right] \qquad (3.3\text{-}4)$$

3.3.2 Assumptions

a) The signal $s^{i_1 \cdots i_N}$ $-\infty < i_k < \infty$ has a known fixed shape.

b) The noise $n^{i_1 \cdots i_N}$ $-\infty < i_k < \infty$ is a random process with unknown distribution.

c) The mean value $E\{n^{i_1 \cdots i_N}\}$ of the noise is known to be equal to zero.

d) The covariance $E\{n^{i_1 \cdots i_N} n^{j_1 \cdots j_N}\}$ of the noise is known.

e) The dimensions are discrete, integer-valued parameters.

f) The observed random process is

$$x^{i_1 \cdots i_N} = s^{i_1 \cdots i_N} + n^{i_1 \cdots i_N}$$

g) The observed random process $x^{i_1 \cdots i_N}$ is convolved with a linear filter with coefficients $f^{i_1 \cdots i_N}$ $1 \le i_k \le m_k$ (to be determined).

h) The actual output of the filter is

$$y^{j_1 \cdots j_N} = \left[f^{i_1 \cdots i_N} , x^{j_1 - i_1 \cdots j_N - i_N} \right]$$

i) The desired output of the filter is $z^{i_1 \cdots i_N}$ $-\infty < i_k < \infty$ where z is a known fixed function.

3.3.3 Statement of Problem

We wish to determine those values of the coefficients $\beta^{i_1 \cdots i_N}$ $1 \le i_k \le m_k$ such that the mean of the sum of the squared-differences between the desired output z and the actual output y is a minimum; that is, such that

$$\alpha = E\{(z-y)^2\} \qquad = \text{minimum} \tag{3.3-5}$$

We define this difference to be

$$\xi^{i_1 \cdots i_N} = z^{i_1 \cdots i_N} - y^{i_1 \cdots i_N}. \tag{3.3-6}$$

3.3.4 Solution

We wish to find values for $\beta^{i_1 \cdots i_N}$ $1 \le i_k \le m_k$ that minimizes α :

$$\alpha = E\{(\xi^{i_1 \cdots i_N})^2\}$$

$$= E\{(z^{i_1 \cdots i_N} - y^{i_1 \cdots i_N})^2\} \tag{3.3-7}$$

$$= E\{(z^{i_1 \cdots i_N} - [\beta^{j_1 \cdots j_N} x^{i_1 \cdots i_N j_1 \cdots j_N}])^2\} \qquad 1 \le j_k \le m_k$$

Taking the derivative, we find

$$\frac{\partial \alpha}{\partial \beta^{i_1 \cdots i_N}} = 0 \Rightarrow E\{x^{i_1 \cdots i_N j_1 \cdots j_N}, \xi^{i_1 \cdots i_N}\} = 0 \tag{3.3-8}$$

Thus, we see that this minimum criterion implies that the error ξ is normal to the process x when dotted with it over the dimensions for which β is to be defined. For this reason, this equation (3.3-8) is known as the normal equation for β.

We can expand this equation to find

$$E\left\{\beta^{k_1 \cdots k_N}\left[x^{i_1-j_1, \cdots, i_N-j_N, k_N}, x^{i_1 \cdots i_N}\right]\right\} = E\left\{x^{i_1-j_1 \cdots i_N-j_N} z^{i_1 \cdots i_N}\right\} \tag{3.3-9}$$

But since

$$x = s + n$$

and

$$E\{s\} = s, \quad E\{z\} = z, \quad E\{n\} = 0$$

we have

$$\left[\beta^{k_1 \cdots k_N}, \left(\left[s^{i_1-j_1, k_1 \cdots i_N-j_N, k_N}, s^{i_1 \cdots i_N}\right] + E\left\{n^{i_1-j_1, k_1 \cdots i_N-j_N, k_N}, n^{i_1 \cdots i_N}\right\}\right)\right] = \left[s^{i_1-j_1 \cdots i_N-j_N}, z^{i_1 \cdots i_N}\right] \tag{3.3-10}$$

or, if we let

$$r^{k_1-j_1 \cdots k_N-j_N} \equiv \left[s^{i_1-j_1, k_1 \cdots i_N-j_N, k_N}, s^{i_1 \cdots i_N}\right] + E\left\{n^{i_1-j_1, k_1 \cdots i_N-j_N, k_N}, n^{i_1 \cdots i_N}\right\} \tag{3.3-11}$$

and

$$g^{j_1 \cdots j_N} \equiv \left[s^{i_1-j_1 \cdots i_N-j_N}, z^{i_1 \cdots i_N}\right] \tag{3.3-12}$$

Then we have

$$\left[\beta^{k_1 \cdots k_N}, r^{k_1-j_1 \cdots k_N-j_N}\right] = g^{j_1 \cdots j_N} \qquad 1 \le j_\ell \le m_\ell \tag{3.3-13}$$

3.3.5 Error Estimation

We can also find a value for α .

$$\alpha = E\left\{\left(z^{i_1 \cdots i_N} - \left[f^{j_1 \cdots j_N}, x^{i_1 - j_1 \cdots i_N - j_N}\right]\right)^2\right\} \qquad (3.3\text{-}14)$$

But, from the normality condition, this is

$$\alpha = E\left\{z^2 - f^{j_1 \cdots j_N}\left[x^{i_1 - j_1 \cdots i_N - j_N}, z^{i_1 \cdots i_N}\right]\right\}$$

$$= [z,z] - \left[f^{j_1 \cdots j_N}, \left[x^{i_1 - j_1 \cdots i_N - j_N}, z^{i_1 \cdots i_N}\right]\right]$$

$$= [z,z] - \left[f^{j_1 \cdots j_N}, g^{j_1 \cdots j_N}\right] \qquad (3.3\text{-}15)$$

4. Recursive Schemes for Normal Equations of Toeplitz Form

4. Recursive Schemes for Normal Equations of the Toeplitz Form.

The solution of the least-square optimum filtering problem as shown in section 3. involves solving a set of simultaneous equations. In general, there will be one equation for each coefficient in the filter. The time and space requirements of a standard simultaneous equation computer routine for such filters is prohibitive for almost all non-trivial problems. This section outlines several more efficient schemes for arriving at the desired filter.

These schemes take advantage of the special forms of the autocorrelation matrix R. In the single and multi-input cases, R has the form

$$\begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{n-1} \\ r_1 & r_0 & r_1 & \cdots & r_{n-2} \\ r_2 & r_1 & r_0 & \cdots & r_{n-3} \\ \vdots & & & & \vdots \\ r_{n-1} & r_{n-2} & r_{n-3} & \cdots & r_0 \end{bmatrix}$$

where $r_i = \overline{r_i}$. In the multi-dimensional case we could formulate a multi-dimensional matrix that would have the same property. That is, in each case, all terms along each diagonal are the same. Thus, given the top row (and possibly the left column if $r_i \neq f(r_i)$) the matrix is fully specified.

The recursive technique involves beginning with a filter of length 1 , using this to find a filter of length 2, etc. At each step of the process, we first find a vector (the prediction error operator) of length m that is normal to each row of the matrix. In the

multiple cases we must find more than one such operator. This operator is then used to increase the length of the filter and of the operator itself.

Once the prediction error operator for an autocorrelation matrix R and the optimum filter for a particular $g = [g_0 \cdots g_m]$ inhomogeneous part of the equation have been found, the optimum filter for which the inhomogeneous part is shifted by one (to, say, $g' = [g_0 \cdots g_{m-1}]$) can be determined by a recursive step.

The principal advantages of using the recursive techniques are time and space savings. The standard solution of simultaneous equations requires time proportional to n^3 and space proportional to n^2. The recursive technique reduces these requirements to n^2 and n for time and space, respectively.

An important side benefit of using this scheme is that we can compute the expected error at each step of the process. This allows us to formulate a criterion for determining the length of the filter. As the filter becomes longer, the expected error will decrease and then level off at some value. The shifting of the desired output (the inhomogeneous part of the equations) may also produce a decrease in expected error.

The development of the recursive scheme for single- and multi-input cases will be made using two different notations. The first notation (labeled Expanded Notation) will be the matrix notation used in section 3. The second notation (labeled Compact Notation) will involve the use of a set of vector operators. Thus, we will define the vectors:

a) $^n r = (r_0, r_1 \cdots, r_n)$ autocorrelation

b) $^n a = (^n a_0, ^n a_1, \ldots ^n a_n)$ prediction error operator

c) $^n b = (^n b_0, ^n b_1, \ldots ^n b_n)$ hindsight error operator

d) $^n f = (^n f_1, ^n f_2, \ldots ^n f_n)$ optimum filter

(The prediction error operator and hindsight error operators will be defined later in this section. The superscript to the left of the symbol indicates the length of the vector as defined above. Thus ${}^{m}r$, ${}^{m}a$, and ${}^{m}b$ are of length $m+1$ and ${}^{m}f$ is of length m). We also define the operators

a) Reversing $R({}^{m}a) = ({}^{m}a_m, {}^{m}a_{m-1}, \ldots {}^{m}a_0)$

b) Zero (increases the length by one)

$$Z({}^{m}a) = ({}^{m}a_0, \ldots, {}^{m}a_m, 0)$$

c) Sliding $D^{i}({}^{m}a) = (r_{-i}, r_{-i+1}, \ldots, r_{-i+m})$

d) Inner Product = Dot Product

$$[{}^{m}a, {}^{m}r] = {}^{m}a_0 r_0 + \cdots + {}^{m}a_M r_m$$

The recursive scheme for the single process was first formulated by N. Levinson (1950). S. M. Simpson, Jr., proposed the recursion for shift of the desired output. E. A. Robinson extended the single process to the multi-input case. Finally, R. A. Wiggins extended the single process to the multi-dimensional case.

4.1 Single Processes

The normal equations for a single process (see section 3.1) optimum filter are of the form

$$\begin{bmatrix} {}^{n}f_1, \ldots, {}^{n}f_M \end{bmatrix} \begin{bmatrix} r_0 & r_1 & \cdots & r_{M-1} \\ r_1 & r_0 & & r_{M-2} \\ \vdots & & \ddots & \vdots \\ r_{M-1} & r_{M-2} & \cdots & r_0 \end{bmatrix} = \begin{bmatrix} g_1 & \cdots & g_M \end{bmatrix} \tag{4.1-1}$$

where r_i, ${}^{n}f_i$ and g_i are scalars. Associated with this equation is the equation of the unit distance prediction error operator ${}^{n}a$ (see section 3.1.7).

$$\begin{bmatrix} {}^{n}a_1, \ldots, {}^{n}a_M \end{bmatrix} \begin{bmatrix} r_0 & \cdots & r_M \\ \vdots & & \\ r_M & \cdots & r_0 \end{bmatrix} = \begin{bmatrix} \alpha, 0, \ldots 0 \end{bmatrix} \tag{4.1-2}$$

where ${}^{n}a_0 = 1$ and α is the expected error (see section 3.1.5). We note that the hindsight error operator ${}^{n}b = [{}^{n}b_M, \ldots, {}^{n}b_0]$, ${}^{n}b_0 = 1$ (i.e., the operator which "predicts" past values of the series from future values) is just the reverse of ${}^{n}a$ since R is symmetric.

4.1.1 The Levinson Recursion to Larger Operators

This development is a modified version of that given by Levinson. He uses the orthogonal operator C (the prediction operator) instead of the prediction error operator.

4.1.1.1 Extension of the Prediction Error Operator

Expanded Notation

The scheme for extending ${}^{n}a$ to ${}^{n+1}a$ is to first make an approximation of ${}^{n+1}a$ by ${}^{n+1}\tilde{a}$ where ${}^{n+1}\tilde{a}_{n+1} = 0$. We substitute this new vector into the equations and examine the solution. Then, by adding a similarly extended hindsight operator to ${}^{n+1}\tilde{a}$ we can get the real solution.

Thus, first we extend ${}^{n}a$ by adding a zero to the right end

$$\left[{}^{n}a_0, \ldots {}^{n}a_n, 0\right] \begin{bmatrix} r_0 & \cdots & r_{n+1} \\ \vdots & & \vdots \\ r_{n+1} & \cdots & r_0 \end{bmatrix} = \left[\alpha_n, 0 \ldots 0, \alpha_n'\right] \tag{4.1-3}$$

where $\alpha_n' = {}^{n}a_0 r_{n+1} + \cdots + {}^{n}a_n r_1$.

Since R is symmetric, the hindsight error operator is just the reverse of ${}^{n}a$. Thus, the extended hindsight operator is the reverse of ${}^{n+1}\tilde{a}$. Thus if we weight and add ${}^{n+1}\tilde{a}$ reversed to ${}^{n+1}\tilde{a}$ and substitute we find

$$\left[{}^{n}a_0, {}^{n}a_1 + k_0 {}^{n}a_n, \ldots, k_n a_0\right] \begin{bmatrix} r_0 & \cdots & r_{n+1} \\ \vdots & & \vdots \\ r_{n+1} & \cdots & r_0 \end{bmatrix} = \left[\alpha_n + k_0 \alpha_n', 0 \ldots 0, \alpha_n' + k_0 \alpha_n\right] \tag{4.1-4}$$

Now, if we choose k_1 such that

$$\alpha_n' + k_1 \alpha_n = 0 \tag{4.1-5}$$

Then we have formed a new ${}^{n+1}a$

$$ {}^{n+1}a = [{}^{n}a_0, \ldots {}^{n}a_n, 0] + k_n [0, {}^{n}a_n, \ldots {}^{n}a_0] \qquad (4.1\text{-}6) $$

Note that the new expected error is

$$ \alpha_{n+1} = \alpha_n + k_n \alpha_n' \qquad (4.1\text{-}7) $$

Compact Notation

The formula corresponding to (4.1-5) is

$$ k_n = - \frac{[{}^{n}a, D^{-n-1}({}^{n}r)]}{\alpha_n} \qquad (4.1\text{-}5a) $$

to (4.1-6) is

$$ {}^{n+1}a = Z({}^{n}a) + k_n R\{Z({}^{n}a)\} \qquad (4.1\text{-}6a) $$

and to (4.1-7) is

$$ \alpha_{n+1} = \alpha_n + k_n [{}^{n}a, D^{-n-1}({}^{n}r)] . \qquad (4.1\text{-}7a) $$

4.1.1.2 Extension of General Filter

Expanded Notation

We now use the prediction error operator ${}^{n}a$ to extend the length of the filter ${}^{n}f$. Here, as in section 4.1.1.1 we make a first approximation to ${}^{n+1}f$ by adding a zero to the end of ${}^{n}f$ to form ${}^{n+1}\bar{f}$.

$$\left[{}^{n}f_1, \ldots, {}^{n}f_n, 0 \right] \begin{bmatrix} r_0 & \cdots & r_n \\ \vdots & \ddots & \vdots \\ r_n & \cdots & r_0 \end{bmatrix} = \left[g_1, \ldots, g_n, \gamma_n \right] \tag{4.1-8}$$

where $\gamma_n = {}^{n}f_1 r_n + \cdots + {}^{n}f_n r_1$

If we weight and add ${}^{n}a$ reverse to ${}^{n+1}\bar{f}$ we get

$$\left[f_1 + k_1 {}^{n}a_n, \ldots, {}^{n}f_n + k_1 {}^{n}a_0, k_1 {}^{n}a_1 \right] \begin{bmatrix} r_0 & \cdots & r_n \\ \vdots & & \\ r_n & \cdots & r_0 \end{bmatrix} = \left[g_0, \ldots, g_n, \gamma_n + k_1 \alpha_n \right] \tag{4.1-9}$$

Now, if $$\gamma_n + k_1 \alpha_n = g_{n+1} \tag{4.1-10}$$

Then the new filter is

$${}^{n+1}f = \left[{}^{n}f_1, \ldots, {}^{n}f_n, 0 \right] + k_1 \left[{}^{n}a_n, \ldots, {}^{n}a_1, {}^{n}a_0 \right] \tag{4.1-11}$$

Compact Notation

The formulae corresponding to (4.1-10) is

$$k_1 = \frac{g_{n+1} - \left[{}^{n}f, D^{-n}({}^{n}r) \right]}{\alpha_n} \tag{4.1-10a}$$

and to (4.1-11) is

$${}^{n+1}f = Z({}^{n}f) + k_1 R({}^{n}a) \tag{4.1-11a}$$

4.1.2 Recursion to Move Output Origin

We wish to find a filter ${}^{n}f'$ that satisfies the normal equation

$$ {}^{n}f' \, R = g' \qquad (4.1\text{-}12) $$

where $g' = [g_0, \ldots g_{n-1}]$ from the solution of the equation

$$ {}^{n}f \, R = g \qquad (4.1\text{-}13) $$

where $g = [g_1 \ldots g_n]$. Note that we could just as easily have chosen to shift g to the left rather than the right.

Our first approximation to the desired filter is to shift the filter ${}^{n}f$ by one.

$$ [0, {}^{n}f_1, \ldots, {}^{n}f_{n-1}] \begin{bmatrix} r_0 & \cdots & r_{n-1} \\ \vdots & & \\ r_{n-1} & \cdots & r_0 \end{bmatrix} = [\gamma_n', g_1 - {}^{n}f_n r_{n-1}, \ldots, g_{n-1} - {}^{n}f_n r_1] \qquad (4.1\text{-}14) $$

where $\gamma_n' = {}^{n}f_1 r_1 + \cdots + {}^{n}f_{n-1} r_{n-1}$

Two types of inhomogeniety have been added to the right side. The first is a weighted version of the ${}^{n}r$ vector. This can be removed by subtracting the negative of the hindsight operator ($[0, {}^{n}a_{n-1}, \ldots {}^{n}a_2, {}^{n}a_1]$).

$$ [0, {}^{n}f_1 - {}^{n}f_n {}^{n}a_{n-1}, \ldots, {}^{n}f_{n-1} - {}^{n}f_n {}^{n}a_1] \begin{bmatrix} r_0 & \cdots & r_{n-1} \\ \vdots & & \\ r_{n-1} & \cdots & r_1 \end{bmatrix} = [\gamma_n'', g_1, \ldots g_{n-1}] \qquad (4.1\text{-}15) $$

where

$$ \gamma_n'' = ({}^{n}f_1 - {}^{n}f_n {}^{n}a_{n-1}) r_1 + \cdots + ({}^{n}f_{n-1} - {}^{n}f_n {}^{n}a_1) r_{n-1} \qquad (4.1\text{-}16) $$

Now we add the prediction error operator to alter to the desired value. Thus, let

$$\gamma_n'' + k_f \cdot \alpha_{n-1} = g. \tag{4.1-17}$$

Then

$$^{n}f' = [0, {}^{n}f_1, \ldots {}^{n}f_{n-1}] - {}^{n}f_n [0, {}^{n-1}a_{n-1}, \ldots {}^{n-1}a_1]$$
$$+ k_f \cdot [{}^{n-1}a_0, {}^{n-1}a_1, \ldots {}^{n-1}a_{n-1}]. \tag{4.1-18}$$

Compact Notation

The formulae corresponding to (4.1-17) is

$$k_f \cdot = g. - \Big[D^{-1}({}^{n}f) - {}^{n}f_n D^{-1}\{R({}^{n-1}a)\}, D^{-1}({}^{n-1}r) \Big]_{\alpha_{n-1}} \tag{4.1-17a}$$

and to (4.1-18) is

$$^{n}f' = D^{-1}({}^{n}f) - {}^{n}f_n D^{-1}\{R({}^{n-1}a)\} + k_f \cdot {}^{n-1}a$$

where $f_0 = {}^{n-1}a = 0$.

4.2 Multi-Input Processes

The normal equations for a multi-input optimum filter (see section 3.2) are of the form

$$[{}^{n}f_1, \ldots, {}^{n}f_n] \begin{bmatrix} r_0 & r_1 & \ldots & r_{n-1} \\ r_{-1} & r_0 & \ldots & r_{n-2} \\ \vdots & & \ddots & \vdots \\ r_{-n+1} & r_{-n+2} & \ldots & r_0 \end{bmatrix} = [g_1, \ldots, g_n] \tag{4.2-1}$$

where $R_i = \overline{R_{-i}}$ is an $n \times n$ matrix
$^n f_i$ is a $\ell \times n$ $\ell \leq n$ matrix
and g_i is a $\ell \times 1$ $\ell \leq n$ matrix.

The unit distance prediction error filter $^n a$ and unit distance hindsight error filter $^n b$ are associated with the R matrix:

$$\left[{}^n a_0, \ldots, {}^n a_n \right] \begin{bmatrix} r_0 & \cdots & r_n \\ \vdots & & \vdots \\ r_{-n} & \cdots & r_0 \end{bmatrix} = \left[\alpha_n, 0 \ldots 0 \right] \tag{4.2-2}$$

$$\left[{}^n b_n, \ldots, {}^n b_0 \right] \begin{bmatrix} r_0 & \cdots & r_n \\ \vdots & & \vdots \\ r_{-n} & \cdots & r_0 \end{bmatrix} = \left[0, \ldots 0, \beta_n \right] \tag{4.2-3}$$

where $^n a_i$, $^n b_i$, α_n and β_n are $n \times n$ matrices,
$^n a_0 = {}^n b_0 = I$, the unit matrix,
and α_n and β_n are the expected error matrices for $^n a$ and $^n b$ respectively.

The only formal difference between this case and the single process case is that the R matrix is no longer symmetric. This causes the hindsight error operator to differ from the prediction error operator reversed.

4.2.1 Recursion to Larger Operators

4.2.1.1 Extension of the Prediction and Hindsight Error Operators

Expanded Notation

We proceed here as in the single process case. The first approximations to the new, longer operators are

made by simply adding a zero to the present operators. Then they are added together to remove the inhomogeneity caused by the zero.

When we add the zero, we find

$$\left[{}^{n}a_0, \ldots, {}^{n}a_n, 0\right] \begin{bmatrix} r_0 & \cdots & r_{n+1} \\ \vdots & & \vdots \\ r_{-n-1} & \cdots & r_0 \end{bmatrix} = \left[\alpha_n, 0 \ldots 0, \alpha_n'\right] \tag{4.2-4}$$

$$\alpha_n' = {}^{n}a_0 r_{n+1} + \cdots + {}^{n}a_n r_1$$

$$\left[0, {}^{n}b_n, \ldots, {}^{n}b_0\right] \begin{bmatrix} r_0 & \cdots & r_{n+1} \\ \vdots & & \vdots \\ r_{-n-1} & \cdots & r_0 \end{bmatrix} = \left[\beta_n', 0 \ldots 0, \beta_n\right] \tag{4.2-5}$$

$$\beta_n' = {}^{n}b_n r_{-1} + \cdots + {}^{n}b_0 r_{-n-1}$$

Now we weight ${}^{n}\tilde{b}$ and add to ${}^{n}\tilde{a}$ and vice versa:

$$\left[{}^{n}a_0, {}^{n}a_1 + k_a {}^{n}b_n, \ldots, k_a {}^{n}b_0\right] \begin{bmatrix} r_0 & \cdots & r_{n+1} \\ \vdots & & \vdots \\ r_{-n-1} & \cdots & r_0 \end{bmatrix} = \left[\alpha_n + k_a \beta_n', 0 \ldots 0, \alpha_n' + k_a \beta_n\right] \tag{4.2-6}$$

$$\left[k_b {}^{n}a_0, {}^{n}b_n + k_b {}^{n}a_1, \ldots, {}^{n}b_0\right] \begin{bmatrix} r_0 & \cdots & r_{n+1} \\ \vdots & & \vdots \\ r_{-n-1} & \cdots & r_0 \end{bmatrix} = \left[\beta_n' + k_b \alpha_n, 0 \ldots 0, \beta_n + k_b \alpha_n'\right] \tag{4.2-7}$$

If we choose k_a and k_b such that

$$\alpha_n' + k_a \beta_n = 0 \tag{4.2-8}$$

and

$$\beta_n' + k_b \alpha_n = 0 \tag{4.2-9}$$

Then the new filters ${}^{m+1}a$ and ${}^{m+1}b$ are

$$ {}^{m+1}a = \left[{}^{m}a_0, \ldots, {}^{m}a_m, 0 \right] + k_a \left[0, {}^{m}b_m, \ldots, {}^{m}b_0 \right] \qquad (4.2\text{-}10) $$

$$ {}^{m+1}b = \left[0, {}^{m}b_m, \ldots, {}^{m}b_0 \right] + k_b \left[{}^{m}a_0, \ldots, {}^{m}a_m, 0 \right] \qquad (4.2\text{-}11) $$

and the new α_{m+1} and β_{m+1} are given by

$$ \alpha_{m+1} = \alpha_m + k_a \beta_m' \qquad (4.2\text{-}12) $$

$$ \beta_{m+1} = \beta_m + k_b \alpha_m' \qquad (4.2\text{-}13) $$

Compact Notation

The compact forms for equations (4.2-8) and (4.2-9) are

$$ k_a = -\left[{}^{m}a, R\left\{ D^{-1}({}^{m}r) \right\} \right] \beta_m^{-1} \qquad (4.2\text{-}8a) $$

$$ k_b = -\left[{}^{m}b, D^{m+1}({}^{m}r) \right] \alpha_m^{-1} \qquad (4.2\text{-}9a) $$

and for equations (4.2-10) and (4.2-11) are

$$ {}^{m+1}a = Z({}^{m}a) + R\left\{ k_a Z({}^{m}b) \right\} \qquad (4.2\text{-}10a) $$

$$ {}^{m+1}b = Z({}^{m}b) + R\left\{ k_b Z({}^{m}a) \right\} \qquad (4.2\text{-}11a) $$

4.2.1.2 Extension of the General Filter

Expanded Notation

We can now use the hindsight error operator to extend the right end of the general filter. The scheme is to add a zero the the operator as a first approximation and then weight and add the hindsight error operator to remove the inhomogeneity. These steps give

$$\left[{}^{n}f_1 + k_f\,{}^{n}b_n, \ldots, {}^{n}f_n + k_f\,{}^{n}b_1, k_f\,{}^{n}b_0\right]\begin{bmatrix} r_0 & \cdots & r_n \\ \vdots & & \vdots \\ r_n & \cdots & r_1 \end{bmatrix} = \left[g_1, \ldots g_n, \gamma_n + k_f \beta_n\right] \tag{4.2-12}$$

where $\gamma_n = {}^{n}f_1 r_n + \cdots + {}^{n}f_n r_1$. Thus, if we set

$$\gamma_n + k_f \beta_n = g_{n+1} \tag{4.2-13}$$

then ${}^{n+1}f$ is given by

$${}^{n+1}f = \left[{}^{n}f_1, \ldots, {}^{n}f_n, 0\right] + k_f\left[{}^{n}b_n, \ldots, {}^{n}b_1, {}^{n}b_0\right] \tag{4.2-14}$$

Compact Notation

The compact forms of equations (4.2-13) and (4.2-14) are

$$k_f = \left(g_{n+1} - \left[{}^{n}f, R\left\{D^{-1}({}^{n}r)\right\}\right]\right)/\beta_n^{-1} \tag{4.2-13a}$$

$${}^{n+1}f = Z\left({}^{n}f\right) + k_f R\left\{{}^{n}b\right\} \tag{4.2-14a}$$

4.2.2 Recursion to Move the Output Origin

The desideratum here is the same as in the single process. From the filter ${}^{n}f$ that satisfies

$$ {}^{n}f \; R = g \tag{4.2-15} $$

where $g = [\, g_0 \cdots g_n \,]$, and the error operators we wish to find ${}^{n}f'$ that satisfies

$$ {}^{n}f' \; R = g' $$

where $$ g' = [\, g_0, \ldots, g_{n-1} \,] \tag{4.2-16} $$

Proceeding as in the single input process, we find that shifting and adding the filter and the hindsight error operator gives

$$ \left[0, {}^{n}f_1 - {}^{n}f_n \, {}^{n-1}b_{n-1}, \ldots {}^{n}f_{n-1} - {}^{n}f_n \, {}^{n-1}b_1 \right] \begin{bmatrix} r_0 & \cdots & r_{n-1} \\ \vdots & \ddots & \vdots \\ r_{n-1} & \cdots & r_0 \end{bmatrix} = \left[\gamma_n'', g_1, \ldots g_{n-1} \right] \tag{4.2-17} $$

where $\gamma_n'' = \left({}^{n}f_1 - {}^{n}f_n \, {}^{n-1}b_{n-1} \right) r_1 + \cdots + \left({}^{n}f_{n-1} - {}^{n}f_n \, {}^{n-1}b_1 \right) r_{n+1}$.

Now solve the equation

$$ \gamma_n'' + k_f \, \alpha_{n-1} = g_0 \tag{4.2-18} $$

to find the weighting for addition of the prediction error operator. The new operator is

$$ \begin{aligned} {}^{n}f' = {} & \left[0, {}^{n}f_1, \ldots, {}^{n}f_{n-1} \right] - \\ & - {}^{n}f_n \left[0, {}^{n-1}b_{n-1}, \ldots, {}^{n-1}b_1 \right] + \\ & + k_f \left[{}^{n-1}a_0, {}^{n-1}a_1, \ldots {}^{n-1}a_{n-1} \right] \end{aligned} \tag{4.2-19} $$

Compact Notation

The compact forms of equations (4.2-18) and (4.2-19) are

$$k_{f'} = \left(g_o - \left[D^{-1}\left({}^{m}f\right) - {}^{m}f_m D^{-1}\left\{ R\left({}^{m-1}b\right)\right\}, D^{-1}\left\{ R\left({}^{m}r\right)\right\} \right]\right) \alpha_{m-1}^{-1} \qquad (4.2\text{-}17a)$$

$${}^{m}f = D^{-1}\left({}^{m}f\right) - {}^{m}f_m D^{-1}\left\{ R\left({}^{m-1}b\right)\right\} + k_{f'}\,{}^{m-1}a \qquad (4.2\text{-}18a)$$

4.3 Multi-Dimensional Processes

In this section we will consider only the 2-dimensional process in order to simplify the notation. All of the techniques given are readily extendable to higher dimensionality.

The 2-dimensional optimum filter is given by the equation (see section 3.3).

$$\left[b^{i_1,i_2},\ r^{i_1-j_1,i_2-j_2} \right] = g^{j_1,j_2} \qquad \begin{matrix} 1 \le j_1 \le m_1 \\ 1 \le j_2 \le m_2 \end{matrix} \qquad (4.3\text{-}1)$$

where m_1 and m_2 are the dimensions of the filter b^{i_1,i_2}. We will seek to extend the filter in the i_2 direction by one unit.

To accomplish this we will need m_2 prediction error filters $a_k^{i_1,i_2}$

$$a_k^{i_1,i_2} = \begin{matrix} 0 & a_k^{i_1,1} & a_k^{i_1,2} & \cdots & a_k^{i_1,m_2} \\ 0 & a_k^{i_2,1} & a_k^{i_2,2} & \cdots & a_k^{i_2,m_2} \\ \vdots & & & & \\ 1 & a_k^{k,1} & a_k^{k,2} & \cdots & a_k^{k,m_2} \\ \vdots & & & & \\ 0 & a_k^{m_1,1} & a_k^{m_1,2} & \cdots & a_k^{m_1,m_2} \end{matrix} \qquad (4.3\text{-}2)$$

where $a_k^{i,0} = \delta(i-k)$,

that satisfy the equations

$$\left[a_k^{i_1,i_2}, r^{i_1-j_1,i_2-j_2} \right] = \alpha_{j_1,k}\, \delta(j_2) \qquad \begin{matrix} 1 \le j_1 \le m_1 \\ 0 \le j_2 \le m_2 \end{matrix} \qquad (4.3\text{-}3)$$

Here, as in the single process, the autocorrelation array has sufficient (centro-) symmetry so that if we reverse the predictors $a_k^{i_1,i_2}$ to become $a_k^{m_1-i_1+1,\, m_2-i_2+1}$ they will still obey the equation.

4.3.1 Recursion to Larger Operators

4.3.1.1 Extension of the Prediction Error Operators

Following the philosophy of the previous sections, we make a first approximation $\tilde{a}_k^{i_1,i_2}$ $\begin{matrix} 1 \le i_1 \le m_1 \\ 0 \le i_2 \le m_2 \end{matrix}$ to the extended prediction error operators $a_k^{i_1,i_2}$ $\begin{matrix} 1 \le i_1 \le m_1 \\ 0 \le i_2 \le m_2+1 \end{matrix}$ by adding a column of zeros after the m_2^{th} column:

$$\left[\tilde{a}_k^{i_1,i_2}, r^{i_1-j_1,i_2-j_2} \right] = \alpha_{j_1,k}\, \delta(j_2) + \beta_{j_1,k}\, \delta(j_2 - m_2 - 1) \qquad \begin{matrix} 1 \le j_1 \le m_1 \\ 0 \le j_2 \le m_2+1 \end{matrix} \qquad ((4.3\text{-}4)$$

Since the r^{i_1,i_2} array is centro-symmetric, reversing both co-ordinates of $\tilde{a}$ has the effect of reversing both co-ordinates of the right hand side:

$$\left[\tilde{a}_k^{m_1-i_1+1,\,m_2-i_2+1},\, r^{i_1-j_1,\,i_2-j_2}\right] = f_{m_1-j_1+1,k}\,\delta(j_2) + \alpha_{m_1-j_1+1,k}\,\delta(j_2-m_2-1) \tag{4.3-5}$$

$$1 \le j_1 \le m_1$$
$$0 \le j_2 \le m_2+1$$

Thus, we can remove the f terms by adding weighted, reversed predictors to the original predictors. That is, if

$$f_{j_1,k} + K_{j_1,k}\,\alpha_{m_1-j_1+1,k} = 0 \tag{4.3-6}$$

Then

$$a_k^{i_1,i_2} = \tilde{a}_k^{i_1,i_2} + \sum_{l=1}^{m_1} K_{k,l}\,\tilde{a}_l^{m_1-i_1+1,\,m_2-i_2+1} \tag{4.3-7}$$

$$1 \le i_1 \le m_1$$
$$0 \le i_2 \le m_2+1$$

and a new error matrix α' is given by

$$\alpha'_{m_1-j_1+1,k} = \alpha_{m_1-j_1+1,k} + \sum_{l=1}^{km} K_{k,l}\,f_{l,k} \tag{4.3-8}$$

4.3.1.2 General Filters

Now we will use the prediction error filters that were found in the preceeding section to extend the length of the general filter. The general filter obeys the relation

$$\left[g^{i_1,i_2},\, r^{i_1-j_1,\,i_2-j_2}\right] = g^{i_1,i_2} \qquad 1 \le j_2 \le m_2 \tag{4.3-9}$$

If we extend the length of g by one in the i_2 direction by adding a column of zeros to form

$$\tilde{f}^{i_1,i_2} = \begin{bmatrix} f^{1,1} & f^{1,2} & \cdots & f^{1,m_2} & 0 \\ f^{2,1} & f^{2,2} & \cdots & f^{2,m_2} & 0 \\ \vdots & & & \vdots & \vdots \\ f^{m_1,1} & f^{m_1,2} & \cdots & f^{m_1,m_2} & 0 \end{bmatrix} \tag{4.3-10}$$

Then the normal equations become

$$\left[\tilde{f}^{i_1,i_2}, r^{i_1-j_1,i_2-j_2} \right] = \gamma^{j_1,j_2} \qquad \begin{matrix} 1 \le j_1 \le m_1 \\ 1 \le j_2 \le m_2+1 \end{matrix} \tag{4.3-11}$$

where

$$\gamma^{j_1,j_2} = \begin{cases} g^{j_1,j_2} & \begin{matrix} 1 \le j_1 \le m_1 \\ 1 \le j_2 \le m_2 \end{matrix} \\ \left[\tilde{f}^{i_1,i_2}, r^{i_1-j_1,i_2-j_2} \right] & \begin{matrix} 1 \le j_1 \le m_1 \\ j_2 = m_2+1 \end{matrix} \end{cases} \tag{4.3-12}$$

The last column of γ can be changed to a given column of g by adding weighted values of the reversed prediction error operators. Thus, if

$$\gamma^{j_1,m_2+1} + K'_{j_1} \alpha_{m_1-j_1+1,k} = g^{j_1,m_2+1} \qquad \begin{matrix} 1 \le j_1 \le m_1 \\ 1 \le k \le m_1 \end{matrix} \tag{4.3-13}$$

then the new f is given by

$$f^{i_1,i_2} = f^{i_1,i_2} + \sum_{k=1}^{m_1} K'_k a_k^{m_1-k+1,\,m_2-k+1} \qquad \begin{matrix} 1 \le i_1 \le m_1 \\ 1 \le i_2 \le m_2 \end{matrix} \tag{4.3-14}$$

The choice of the end of f to extend was quite arbitrary. We could have chosen to form a new f^{i_1,i_2} $\begin{matrix} 1 \le i_1 \le m_1 \\ 0 \le i_2 \le m_2 \end{matrix}$ by adding the column of zeros to the other end. For this case, we would then use the unreversed prediction error filters and a slightly altered form of the α matrix.

In the same manner, the i_1 dimension of f can be extended if we are given m_1 prediction error filters in that direction. The development proceeds exactly as that shown above.

5. The Wiener-Masani Technique of Multiple Spectral Factorization

THE WIENER-MASANI TECHNIQUE OF MULTIPLE SPECTRAL FACTORIZATION

Abstract

The prediction problem for single stationary time series is reviewed and the least square and Kolmogoroff solutions given. Extension is then made to the multiple case, the least squares equations set up and the Wiener-Masani factorization described. Heuristic use is made of the Hilbert space property of time series. A digital computer program for performing the Wiener-Masani factorization is discussed.

1.1 Introduction

The problem of prediction (extrapolation) of a stationary stochastic process has been under intensive study for over 20 years. Since the earliest work of Wiener, Wold, Kolmogoroff and others, a considerable literature has developed in the theory and techniques of single time series, be it to predict it an arbitrary distance into the future, to interpolate it, or solve one of the manifold filter problems. The theory of single time series may be said to be nearly complete.

The problems inherent in handling more than one time series simultaneously are, however, far from solved, and the mathematical apparatus necessary is in many ways cumbersome. Since the case of the single time series has been so well resolved, one might question the necessity of treating the multiple case. The reason is, of course, that we are interested in a group of time

series where the members are related to each other, i.e., correlated.

An example of such a group where information is contained in the cross-correlations is the 3 components of motion of a seismograph. Clearly the motion in a given direction will bear some relation to that in a perpendicular direction. The information contained in the cross-correlations ought to be of aid in predicting or filtering such a process.

In section 3.2 of this report, a method of handling multiple prediction by least squares techniques is given. An alternative method (due to Wiener and Masani) will be discussed here from a basic approach. The method relies on very abstract geometric properties of time series. The use of this method will help illuminate these properties. In addition, since the method is computationally feasible, we will achieve a useful check on the least squares approach.

Since much of what we will do has a close analogue in the well-understood scalar series case, a brief review of the prediction-factorization problem will be helpful to clarify the analogies, and the difficulties, that arise in higher dimensions.

The discussion will be confined to the discrete case. The approach will be intuitive and heuristic as rigorous documentation is amply available in the literature.

2. Single Time Series

In what has become known as the Wiener-Kolmogoroff theory, we consider a stationary time series X_t, and for the most part deal directly with two quantities, the correlation

$$r_\tau = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{+N} X_t X_{t+\tau} \qquad (1)$$

and the power density spectrum

$$\Phi(\omega) = \sum_{\tau=-\infty}^{\infty} r_\tau e^{i\omega\tau}$$

r_τ and $\Phi(\omega)$ are clearly Fourier transforms of each other. In the most general case, the function Φ will not exist as such. Instead, it becomes necessary to use the spectral distribution function $\lambda(\omega)$, where the relation to the correlation is

$$r_\tau = \frac{1}{2\pi} \int e^{-i\omega\tau} d\lambda(\omega)$$

the integral being taken in a Riemann-Stieltjes sense. If λ is absolutely continuous then, Φ exists, and

$$\Phi(\omega) = \frac{d\lambda(\omega)}{d\omega} .$$

In circumstances where Φ does not exist, the theory is more complex; as Wiener (1950, p. 58) points out, however, in practice the behavior of nature is not such as to give pure spikes in a spectrum, nor to give pure jumps at neighboring points--corresponding to a non-absolutely continuous spectral distribution. Later, we will give an explicit criterion for the existence of an absolutely continuous distribution.

Let the entire past of a time series be given, i.e. at time t, the set of numbers $(X_t, X_{t-1}, X_{t-2}, \ldots)$ is known. We wish to predict the value of X_t at time $t+\alpha$.

At best this can be done with some smallest error. There will always be an error, since by definition no random process can be exactly predicted. We thus seek an $\hat{X}_{t+\alpha}$ such that the difference $(\hat{X}_{t+\alpha} - X_{t+\alpha})^2$ is as small as possible.

The procedure (Levinson, 1950) is to assume that $\hat{X}_{t+\alpha}$ is given by some linear combination of the past of the time series, i.e.

$$\hat{X}_{t+\alpha} = \sum_{K=\alpha}^{N} a_K X_{t+\alpha-K}$$

where N is made larger for a better fit. Note that it suffices to consider $\alpha = 1$. Any other distance can be determined similarly. Note, too, that stationarity insures that a_K will be the same no matter what choice of origin is made.[1]

The square error is

$$\epsilon_t^2 = (\hat{X}_t - X_t)^2 = \left(\sum_{K=1}^{N} a_K X_{t-K} - X_t\right)^2$$

and its mean over the times -T to + T is

$$\overline{\epsilon_T^2} = \frac{1}{2T+1}\sum_{t=-T}^{+T} \epsilon_t^2 = \frac{1}{2T+1}\sum_{t=-T}^{+T}\left(\sum_{K=1}^{N} a_K X_{t-K} - X_t\right)^2$$

To minimize, differentiate with respect to a_K and set the derivatives equal to zero.

---

[1] By stationary is meant, loosely, that the underlying probabilistic structure is independent of time.

Thus

$$0 = \frac{1}{2T+1}(2)\sum_{t=-T}^{+T}\sum_{K=1}^{N}\left(a_K X_{t-K} X_{t-r} - X_t X_{t-r}\right)$$

or

$$\sum_{K=1}^{N}\frac{1}{2T+1}\sum_{t=-T}^{+T} a_K X_{t-K} X_{t-r} = \frac{1}{2T+1}\sum_{t=-T}^{+T} X_t \; X_{t-r} \qquad (2)$$

But $\lim_{T\to\infty}\frac{1}{2T+1}\sum_{t=-T}^{+T} X_t \; X_{t+J} = r_J$, the autocorrelation. Hence eqn. (2) becomes

$$\sum_{K=1}^{N} a_K r_{K-r} = r_J \qquad (3)$$

These equations are discussed in section 3.1 of this report.

Note that we could have obtained the same result directly by requiring that the error ϵ_t be uncorrelated with the time series:

$$\frac{1}{2T+1}\sum_{t=-T}^{+T}\epsilon_t X_{t-r} = 0 = \frac{1}{2T+1}\sum_{t=-T}^{+T}\sum_{K=1}^{N} a_K X_{t-K} X_{t-r} - \frac{1}{2T+1}\sum_{t=-T}^{+T} X_t X_{t-r}$$

or

$$\sum_{K=1}^{N} a_K r_{K-r} = r_r$$

The identity of the two approaches is significant and we will return to this point later.

In the above prediction, no estimate of the error involved was obtained. To find an estimate, a deeper knowledge of the nature of time series is necessary.

The cornerstone of this knowledge is a theorem of Wold (Doob, 1953, p. 576).

Th. 1. (Wold Decomposition) A stationary stochastic process X_t may be uniquely represented as the sum of two processes, $X_t = U_t + V_t$

where $U_t = \sum_{K=0}^{\infty} a_K \vartheta_{t-K}$

such that $a_0, a_1, a_2, \cdots$ is minimum-delay

where $\sum_{K=0}^{\infty} a_K^2 < \infty$, $a_0 > 0$,

$E(\vartheta_r \vartheta_s) = \delta_{rs}$ (4) and $E(\vartheta_t) = 0$, (5).

V_n is deterministic and $E(U_t, V_r) = 0$.

By "deterministic" is meant, in a loose sense, that a knowledge of the zeroes of V_t in some finite interval is sufficient to determine its zeroes in any other interval. It is just the absence of such a process V_t in X_t that will insure that the spectral distribution function is absolutely continuous.

For the present, it will be assumed that

$$X_t = \sum_{K=0}^{\infty} a_K \vartheta_{t-K} \tag{6}$$

The transient, or "wavelet" has "minimum phase characteristic" (Robinson, 1954, 1962) i.e. for all other transients b_K into which we could decompose X_t, a_K has the property $\sum_{K=0}^{N} a_K^2 \geq \sum_{K=0}^{N} b_K^2$, $0 \leq N < \infty$.

a_K has an additional, very important, property. If an inverse wavelet d_K is defined such that

$$\sum_{K=0}^{\infty} d_K a_{l-K} = \begin{cases} 0, & l \neq 0 \\ 1, & l = 0 \end{cases}$$

then $\sum_K d_K^2 < \infty$, i.e. the inverse wavelet

is stable and eqn. (6) can be inverted to give

$$\vartheta_t = \sum_{k=0}^{\infty} \delta_k X_{t-k}$$

a_k is the only wavelet with this property.

The prediction problem, in the light of the Wold Decomposition, can be given a more explicit form. Suppose an estimate at time t of $X_{t+\alpha}$ is wanted. At $t = t+\alpha$,

$$X_{t+\alpha} = \sum_{t=0}^{\infty} a_k \vartheta_{t+\alpha-k}$$

At time t, the terms $a_0 \vartheta_{t+\alpha}, a_1 \vartheta_{t+\alpha-1}, \cdots a_\alpha \vartheta_t$ are unknown. The best prediction is thus based on the known terms and

$$\hat{X}_{t+\alpha} = \sum_{t=\alpha}^{\infty} a_k \vartheta_{t+\alpha+k}$$

where the missing terms are a measure of the error.

The expected mean square error is

$$E(\hat{X}_{t+\alpha} - X_{t+\alpha})^2 = E(a_0 \vartheta_{t+\alpha} + \alpha_2 \vartheta_{t+\alpha-1} + \cdots + a_{\alpha-1} \vartheta_{t+1})^2$$

But by (4) this is $\sum_{m=0}^{\alpha-1} a_t^2$

and we have an explicit expression for the expected error.

In order to actually make the prediction, the ϑ_t must be known. Hence a solution to the inverse problem

$$\vartheta_t = \sum_{k=0}^{\infty} \delta_k X_{t-k}$$

is sought. Since δ_k is inverse to a_k , a knowledge of a_k will suffice.

The autocorrelation of X_t is

$$\lim_{T\to\infty} \frac{1}{2T+1} \sum_t \left[\left(\sum_K a_K g_{t-K}\right)\left(\sum_P a_P g_{t+\tau-P}\right)\right]$$

$$= \lim_{T\to\infty} \frac{1}{2T+1} \sum_K \sum_P a_K a_P \sum_t g_{t-K} g_{t+\tau-P}$$

$$= \lim_{T\to\infty} \frac{1}{2T+1} \sum_K \sum_P a_K a_P \delta_{P-\tau,K} = \lim_{T\to\infty} \frac{1}{2T+1} \sum_P a_{P-\tau} a_P$$

which is the autocorrelation of a_t alone. Hence

$$\Phi(\omega) = \sum_{\tau=-\infty}^{\infty} r_\tau e^{i\omega\tau}$$

and depends only on a_K . An alternative means of computing $\Phi(\omega)$ is to take the Fourier transform of $a = A(\omega)$ and

$$\Phi(\omega) = A(\omega) A^*(\omega) \tag{7}$$

All phase information about a_t is lost in computing $A(\omega) A^*(\omega)$, and there are an infinite number of wavelet transforms $B(\omega)$, such that $B(\omega) B^*(\omega) = A(\omega) A^*(\omega)$. In order to find $A(\omega)$ from the spectrum (whose transform is the desired wavelet) use must be made of the unique property of a_K , -- it is the only invertible wavelet.

Theorem 2. (Jensen's Formula, Ahlfors, 1953, p. 185)

Consider an analytic function $f(z)$ and suppose $f(z) \neq 0$, $|z| \leq 1$. Then in the region $|z| \leq 1$, $\log |f(z)|$ satisfies Laplace's equation and by the standard theorems of complex variables

$$\log |f(0)| = \frac{1}{2\pi} \int_0^{2\pi} \log |f(e^{i\theta})| \, d\theta \tag{8}$$

(Cauchy integral formula). If there are zeroes of f(z)

inside $|z| \leq 1$, then equation (8) must be modified. Suppose the zeroes are located at $\zeta_i, i = 1, \cdots, n$, where n includes multiplicities. Denote their complex conjugates by $\overline{\zeta_i}$. The quantity

$$Q(z) = \prod_{i=1}^{n} \frac{1 - \overline{\zeta_i} z}{(z - \zeta_i)}$$

has poles at the zeroes of $f(z)$ and $Q(e^{i\theta}) = 1$.

$F(z) = Q(z) f(z)$ is analytic for $|z| \leq 1$

and by (8)

$$\log |F(0)| = \frac{1}{2\pi} \int_0^{2\pi} \log |f(e^{i\theta})| \, d\theta$$

$$= \log |f(0)| + \log Q(z)$$

$$= \log f(0) + \sum_{i=1}^{n} \log \frac{1}{|\zeta_i|} = \frac{1}{2\pi} \int_0^{2\pi} \log |f(e^{i\theta})| \, d\theta$$

Hence,

$$\log |f(0)| = -\sum_{i=1}^{n} \log \frac{1}{|\zeta_i|} + \frac{1}{2\pi} \int_0^{2\pi} \log |f(e^{i\theta})| \, d\theta \qquad (9)$$

This is sometimes also expressed as an inequality,

$$\log |f(0)| \leq \frac{1}{2\pi} \int_0^{2\pi} \log |f(e^{i\theta})| \, d\theta \qquad (10)$$

where equality holds if $|\zeta_i| > 1$ (the condition is sufficient, but not necessary).

The representation is unique is the sense that no other function $g(z)$ has the property that

$$\lim_{\rho \to 1} g(\rho e^{i\theta}) \to f(e^{i\theta})$$

Suppose Φ is factored in some arbitrary way

$$\Phi(e^{i\theta}) = F(e^{i\theta}) F^{*}(e^{i\theta})$$

By Jensen's formula,

$$\log |F(0)|^2 \leq \frac{1}{2\pi} \int_0^{2\pi} \log |F(e^{i\theta})|^2 \, d\theta$$

if $F(z)$ has no zeroes $|z| \leq 1$, then strict equality holds and

$$|F(0)|^2 = e^{\frac{1}{2\pi} \int_0^{2\pi} \log |\Phi(e^{i\theta})| \, d\theta} \tag{11}$$

The absence of zeroes of $F(z)$ implies that $\frac{1}{F(z)} = G(z)$ is analytic $|z| \leq 1$ and the Fourier transform of $G = g_t$ exists. By the convolution theorem of Fourier transforms, g_t is inverse to the Fourier transform f_t of F:

$$\sum_k f_k \, g_{t-k} = \begin{cases} 0 & t \neq 0 \\ 1 & t = 0 \end{cases}$$

Hence f_t is invertible and must be just the wavelet of the Wold Decomposition. Therefore, we have proved the following theorem:

Theorem 3.

If $\Phi(e^{i\theta}) = A(e^{i\theta}) \, A^*(e^{i\theta})$, then

(a) $|A(0)|^2 \leq e^{\frac{1}{2\pi} \int_0^{2\pi} \log \Phi(e^{i\theta}) \, d\theta}$

(b) Equality holds if and only if $A(\omega)$ is the minimum phase wavelet of the Wold Decomposition.

We now have sufficient information about $A(\omega)$ to generate it.

From the uniqueness part of Jensen's formula, $A(z)$ is analytic $|z| \leq 1$, and $|A(0)|^2 = e^{\frac{1}{2\pi} \int_0^{2\pi} \log \Phi(e^{i\theta}) \, d\theta}$, such that $\lim_{\rho \to 1} |A(\rho e^{i\theta})|^2 = \Phi(e^{i\theta})$

Therefore, (Robinson, 1954, p. 125) let us seek a power series expansion of $A(z)^2$ such that

$$A(z)^2 = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots$$

and a Fourier series expansion of $\log \Phi(e^{i\theta})$

$$\log \Phi(e^{i\theta}) = \sum_{n=0}^{\infty} \beta_n \cos n\theta$$

Hence for $|z| \leq 1$,

$$\log |A(z)|^2 = \sum_{n=0}^{\infty} \beta_n z^n$$

or

$$|A(z)|^2 = e^{\sum_{n=0}^{\infty} \beta_n t^n} \tag{12}$$

Thus, the procedure is to expand $\log \Phi(e^{i\theta})$ in a cosine series and solve for the power series coefficients of $|A(z)|^2$ in (12), and

$$|A(0)|^2 = e^{\beta_0} = e^{\frac{1}{2\pi}\int_0^{2\pi} \log \Phi(e^{i\theta})\, d\theta}$$

This is called the Kolmogoroff factorization procedure.

3. Multiple Time Series

3.1 Preliminary Considerations

Notation: The "order" of a time series will refer to the number of single time series which make it up. Thus, the single time series of the previous sections have order 1 (scalar case).

The "dimension" of the time series will refer to its physical display. Thus, a single time series has dimension 1. Four single time series displayed as a square array have order 4 and dimension 2, etc.

Quantities which are matrices or vectors, and which may be confused with corresponding scalar quantities are barred: X_t is one time series; $\overline{X}_t$ is of dimension 2 or more, etc.

Outwardly many of the techniques for handling multiple time series are the same as in the scalar case.

We will make use of this similarity as far as possible in what follows. There is as yet no complete solution to the problem of multiple prediction, and there are at least two related, but different, approaches to the problem (Wiener and Masani, 1957, 1958; Helson and Lowdenslager, 1958).

This paper will discuss the Wiener-Masani technique, as it lends itself to an intuitive approach. The two papers of Wiener and Masani are clear and elegant; at the risk of oversimplification, we will, therefore, restrict this paper to a purely heuristic treatment.

Restriction will be made to 2-dimensional processes of order n, i.e. n-scalar time series written as a vector $\overline{X}_t$, where the ith scalar series is denoted X_t^i. This is no restriction on the mathematics and avoids the cumbersome notation necessary in handling arrays of time series.

Definition: A function $\overline{F}(\theta)$ will be said to be $n \times m$ matrix valued if $\overline{F}(\theta)$ is given by an $n \times m$ matrix.

Definition: By $G(\theta) = \int_0^\theta \overline{F}(\theta)\, d\theta$ will be meant the $n \times m$ matrix whose elements are the integrals of the corresponding elements of $\overline{F}(\theta)$.

Consider a set of matrix valued functions $\{\overline{F}\}$ such that

$$\int_0^\Lambda |\overline{F}(\gamma)|^2 d\gamma < \infty$$

Then the set $\{\overline{F}\}$ is said to be in L_2 over $[0, A]$.

Suppose $\overline{F}, \overline{G} \in L_2$. Then define the "inner product" of $\overline{F}$ and $\overline{G}$ as

$$\frac{1}{A}\int_0^A \text{trace}\left[\overline{F}(y)\,\overline{G}^*(y)\right] dy \tag{13}$$

which will be denoted $((\overline{F}, \overline{G}))$.

Define the "norm" of a function $\overline{F} \in L_2$ as the square root of its inner product with itself, written

$$\|\overline{F}\| = \sqrt{((\overline{F}, \overline{F}))} = \left[\frac{1}{A}\int_0^A \overline{F}(y)\,\overline{F}^*(y)\, dy\right]^{\frac{1}{2}} \tag{14}$$

Notice that the inner product can be interpreted as a dot product--it gives a measure of the "overlap" of $\overline{F}$ and $\overline{G}$ over their region of definition--the interval $[0, A]$.

The norm of F is interpretable as the "length" of F. It has all the properties normally associated with a length--i.e.

$$\|\overline{F}\| > 0, \quad \overline{F} \neq 0 \qquad (a)$$
$$\|a\overline{F}\| = |a|\,\|\overline{F}\| \qquad (b) \tag{15}$$

The "distance" between 2 functions $\overline{F}$ and $\overline{G}$ is

$$\|\overline{F} - \overline{G}\| > 0, \quad \overline{F} \neq \overline{G} \tag{15a}$$

The triangle inequality exists:

$$\|\overline{F} - \overline{G}\| \leq \|\overline{F} - \overline{H}\| + \|\overline{H} - \overline{G}\| \tag{15b}$$

Using the definition of inner product, we can speak of 2 functions as being orthogonal if and only if $((\overline{F}, \overline{G})) = 0$.

If the functions $\{\overline{F}_i\}$ obey these rules, and if every Cauchy sequence in L_2 converges, then the space L_2 is called a Hilbert space (Halmos, 1951, p. 17).

Notice that in such a Hilbert space, given a sequence of non-zero functions $\overline{F}_i$ an orthonormal sequence can be generated, i.e. if

$$\| \overline{F}_i \| = a_i$$

then $\overline{G}_i = \frac{1}{\sqrt{a_i}} \overline{F}_i$ will have "length" 1 and the Gram-Schmidt process with the definition (13) can be used to construct the next function.

Suppose two time series $\overline{X}_r$ and $\overline{Y}_s$ of order n are taken as $n \times 1$ matrix valued functions where X_r^i and Y_s^j are each in L_2. Define the matrix,[1]

$$\overline{H} = (\overline{X}_r, \overline{Y}_s)$$

where

$$H_{r,s}^{i,j} = \lim_{T \to \infty} \int_{-T}^{+T} X_r^i Y_s^j \, dt \qquad (16)$$

This term is just the cross-correlation between X^i and Y^j. The matrix $\overline{H}$ is called the Gramian of $\overline{X}$ and $\overline{Y}$. In time series analysis, it is a cross-correlation matrix, where $\overline{H} = \overline{H}^*$. If, as here $\overline{X}$ and $\overline{Y}$ are stationary processes, $\overline{H}$ depends only on $r - s$, not on r and s individually.

The inner product (13) then becomes

$$((X_r, Y_s)) = \text{trace } \overline{H}_{r-s}$$

---

[1] The integral should really be defined in terms of an underlying probability space, but the effective result is the same as eqn. (16).

In the scalar theory of time series, a fundamental role was played by the correlation. Suppose we attempt to use the Gramian matrices in an analogous way. Instead of taking the inner product of $\bar{X}, \bar{y}$ to be *trace* $\bar{H}$, let the inner product be given by $\bar{H}$ itself.

Then $\bar{X}_r$ will be said to be orthogonal to $\bar{y}_s$ if, and only if, $\bar{H}_{r-s} \equiv 0$, and that $\bar{X}_r$ has unit length if, and only if, $(\bar{X}_r, \bar{X}_r) = I$.

This new space, to be denoted $\mathcal{L}_2$, is no longer a Hilbert space. It differs from a Hilbert space in a very profound fashion. It no longer satisfies (15) or (15a). Moreover, consider the generation of an orthonormal sequence of functions $\bar{f}_i$.

Suppose $(\bar{f}_i, \bar{f}_i) = \bar{K}$. Then presumably the quantity $\bar{f}_i = \sqrt{\bar{K}^{-1}}\, f_i$ [1] will have unit length. But $\bar{K}$ is a matrix--it may not be invertible, and, if it is not, a normalized sequence cannot be generated, though an orthogonal sequence may be, of the form such that

$$(\bar{X}_m, \bar{X}_n) = \delta_{m,n} \bar{K}$$

The underlying distinction between the spaces $\mathcal{L}_2$ and L_2, is that L_2 is a Hilbert space, whereas $\mathcal{L}_2$ is the Cartesian product of a set of Hilbert spaces. Each X_r^i individually satisfies the requirements of a function in L_2. In creating $\mathcal{L}_2$ we have adjoined the X_t^i in groups of n. This is the relationship and the

---

[1] For the definition of square root of a matrix see Appendix A.

distinction.

The difficulty with singular matrices is one of the problems of the multiple theory.

Despite this difficulty, it is still convenient to think of $\mathcal{L}_2$ as defining a linear vector space. The visualization of a time series as a vector in a linear vector space is a very useful concept.

The past of a time series $\overline{X}_t$ can be thought of as "spanning" a subspace of $\mathcal{L}_2$.

Let M_n denote the subspace spanned by X_n and its past. At a later time t, the series and its past spans a new subspace M_t where $M_t \supset M_n$, and we are able to speak of the difference space $M_t - M_n$. This concept can be given a rigorous formulation.

In this same sense one could speak of unit vectors $\{\bar{e}_t\}$ in $\mathcal{L}_2$, much as we use unit vectors in ordinary Euclidean 3-space or use eigenvalues as "unit vectors".

If a complete set $\{\bar{e}_t\}$ could be found, any time series $\overline{X}_t$ could be expanded as

$$\overline{X}_t = \sum_{k=0}^{\infty} \bar{a}_k \bar{e}_{t-k}$$

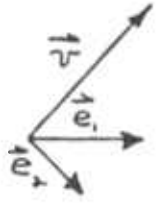
However, as we have pointed out above, it may be that normalization is impossible. This simply implies that though we may seek a complete set, we may not be able to normalize it.

In summary, for conceptual purposes, a time series will be regarded as defining a linear vector space,

$\mathcal{L}_2$, with inner product given by the Gramian, with $\bar{X} \perp \bar{y}$ if, and only if, $(\bar{X}, \bar{y}) = 0$. As time increases, a series spans a larger and larger region of $\mathcal{L}_2$, i.e. a new "basis vector" is added with each time interval. $\mathcal{L}_2$, of course, has the property that the number of its dimensions is infinite and the even more unusual property that non-zero vectors may not be normalizeable. These two properties are a reminder that the picture of a time series as a vector is purely for purposes of intuitive argument, and has only a very rough truth.

Consider the prediction problem once again. Given $\overline{X}_t$ its entire past, and its Gramian, we wish to predict in the best possible way $\overline{X}_{t+\alpha}$. In terms of the vector space picture, what we have in $\overline{X}_t$ and its past is a portion, M_t of $\mathcal{L}_2$. In order to know $\overline{X}_{t+\alpha}$, we would have to know $M_{t+\alpha}$. The question is how to best represent $M_{t+\alpha}$ with a knowledge of M_t only, lacking $M_{t+\alpha} - M_t$.

In ordinary Euclidean 3-space, we could think, in analogy, of the problem of how to best represent a 3-dimensional vector with only 2 unit vectors, i.e.



The best representation of $\vec{v}$ in terms of $\vec{e}_1$, and $\vec{e}_2$ will necessarily leave as error a vector that is perpendicular to both $\vec{e}_1$, and $\vec{e}_2$. To find the best representation we seek constants a and b such that $\hat{\vec{v}} = a\vec{e}_1 + b\vec{e}_2$ and $\hat{\vec{v}} - \vec{v}$ is perpendicular to both $\vec{e}_1$ and $\vec{e}_2$. This gives

$$\hat{\vec{v}} = (\vec{v} \cdot \vec{e}_1)\vec{e}_1 + (\vec{v} \cdot \vec{e}_2)\vec{e}_2 .$$

To represent $\bar{X}_{t+\alpha}$ in terms of its past, we seek matrices $\bar{a}_\kappa$ such that $\hat{\bar{X}}_{t+\alpha} = \sum_{\kappa=0}^{\infty} \bar{a}_\kappa \bar{X}_{t-\kappa}$

where $\hat{\bar{X}}_{t+\alpha} - \bar{X}_{t+\alpha}$ is to be perpendicular to $\bar{X}_t$ $t=0,\cdots,-\infty$. (17)

The effect of using matrix coefficients is to make use of the past of all the time series to represent any one of them--thus increasing the number of basis vectors.

Using (17) as a condition and the inner product, (13) we have

$$\left(\hat{\bar{X}}_{t+\alpha} - \bar{X}_{t+\alpha}, \bar{X}_{t-r}\right) = 0 \qquad r = 0, \cdots, \infty$$

$$\left(\sum_\kappa \bar{a}_\kappa \bar{X}_{t-\kappa}, \bar{X}_{t-r}\right) - \left(\bar{X}_{t+\alpha}, \bar{X}_{t-r}\right) = 0$$

or

$$\sum_\kappa \bar{a}_\kappa \bar{R}_{\kappa-r} = \bar{R}_{\alpha+r} \qquad (18)$$

Eqn. (15) will be recognized as a multiple version of eqn. (3) for the scalar case. It was pointed out above that the requirement that the prediction error be uncorrelated with the time series is equivalent to minimizing the mean square error. This can also be shown to be the case here (see Section 3.2 of this report).

In practice, of course, the entire past of the series is not known and $r = 0, \cdots, N$, some N.

3.2 Wiener-Masani Technique

In considering a scalar time series, we sought another formulation of the prediction problem--in terms of the Wold Decomposition and the Kolmogoroff factorization technique to gain insight into the nature of the time series process and to determine the error involved.

A solution completely analogous to the Kolmogoroff spectral factorization is not known. The Wiener-Masani technique, however, provides a similar exact factorization--though the method of solution is quite distinct. It is more complex than the Kolmogoroff solution due to the matrix nature of the quantities involved. A first step in this technique is the determination of a generalized version of the Wold Decomposition.

Consider $\overline{X}_n$ and its past M_{n-1}. The projection of $\overline{X}_n$ on its past we will denote as $(\overline{X}_n \mid M_{n-1})$, where the projection can be interpreted as a vector projection--onto a plane, for example.

If $\overline{X}_n - (\overline{X}_n \mid M_{n-1}) \neq 0$, we will say, following Wiener and Masani, that the time series is non-deterministic and define $\overline{\vartheta}_n = \overline{X}_n - (\overline{X}_n \mid M_{n-1})$. (19)

For obvious reasons, $\mathcal{I}_n$ is called the "innovation" process of $\bar{X}_n$.

The Gramian of $\mathcal{I}_n$ is

$$(\mathcal{I}_m, \mathcal{I}_n) = \delta_{mn} G, \quad G = (\mathcal{I}_0, \mathcal{I}_0)$$

The vector interpretation of this is that the innovations must be orthogonal to each other--by definition.

The "rank" of the process is the rank of G. If G^{-1} exists, then the process is said to be of full rank and we can normalize the $\mathcal{I}_n$. Here is a demonstration of the critical importance of the matrix nature of the inner product of $\mathcal{L}_2$. The prediction problem for processes where G is of less than full rank has not been solved (Helson & Lowdenslager, 1961, discuss this problem).

The important point is that for full rank processes a Wold Decomposition exists. A decomposition for processes of less than full rank is not known. Therefore, in the remainder of this paper, unless otherwise stated, G will be assumed to have full rank.

Theorem (4). (Wold Decomposition. See Wiener and Masani, 1958, p. 137 for a proof).

If $\mathcal{I}_n$ is the innovation process of a non-deterministic time series of order q, then

(a) $\bar{X}_n = \bar{U}_n + \bar{V}_n$ such that

$$\bar{U}_n = \sum_{K=0}^{\infty} \bar{a}_K \mathcal{I}_{n-K} , \quad \bar{U}_n \perp \bar{V}_n$$

and $\|\bar{u}_n\|^2 = \sum_{\kappa=0}^{\infty} ((\bar{a}_\kappa G^{\frac{1}{2}}))^2 < \infty$

where $G = (\vartheta_0, \vartheta_0)$. The matrices $\bar{a}_\kappa$ are the Fourier coefficients of $\bar{u}_n$ with respect to the α_κ, the $\bar{a}_\kappa$ are minimum-delay, and $\bar{v}_n$ is deterministic.

Note that this is a complete matrix analogue to theorem 1. As in theorem 1, it is required that there be an inverse matrix wavelet α_κ, such that $\sum_{\kappa=0}^{\infty} ((\bar{a}_\kappa (G^{-1})^{\frac{1}{2}}))^2 < \infty$

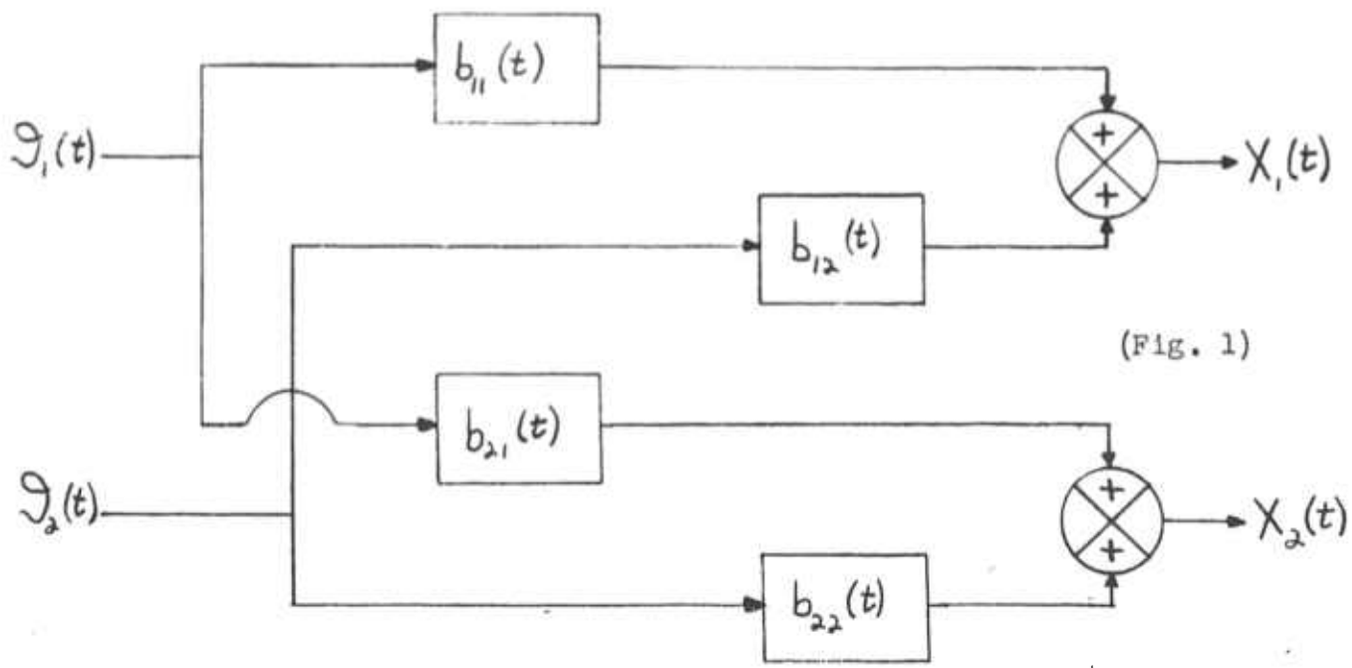
The matrices $\bar{a}_\kappa$ are determined by

$$(\bar{u}_n, \vartheta_r) = \left(\sum_{\kappa=0}^{\infty} \bar{a}_\kappa \vartheta_{n-\kappa}, \vartheta_r\right) = \bar{a}_\kappa \delta_{r,n-\kappa} G$$

$$= \bar{a}_{n-r} G = (u_0, \vartheta_{n-r}) \quad \text{by stationarity.}$$

or $\bar{a}_\kappa G = (u_0, \vartheta_{-\kappa}) = (X_0, \vartheta_{-\kappa})$.

For a time series of order 2, a picture of the structure of such a process is given by Robinson (1962, p. 94).



(Fig. 1)

Since G is invertible, we could normalize $\mathcal{Y}_n$ to a sequence $\mathcal{E}_n$, suppressing all mention of the G matrix. Since it is sometimes convenient not to normalize, we will use the $\mathcal{Y}_n$ sequence.

Define the past of the $\bar{v}_k$ process by η_k. Then, in the vector space analogy, the meaning of $\bar{v}_k$ is that $\lim_{n\to-\infty} \eta_n \neq \{0\}$ and $\eta_{-\infty} = \eta_k$ all k. Thus, the "remote past" of v_k is some finite volume in $\mathcal{L}_2$ and does not change in time.

If a process has $M_{-\infty} = \{0\}$, then it is said to be "regular." All processes of the form of $\bar{u}_k$ are regular. Such processes will have absolutely continuous spectral distribution matrices. We will discuss only these processes; i.e., from now on $\bar{x}_k$ is assumed to be of the form

$$\bar{x}_t = \sum_{k=0}^{\infty} \bar{a}_k \mathcal{Y}_{t-k} \tag{20}$$

The prediction problem in these terms is the problem of finding the best representation of

$$\bar{x}_{t+\alpha} = \sum_{k=0}^{\infty} \bar{a}_k \mathcal{Y}_{t+\alpha-k}$$

where we do not know the terms $\bar{a}_0 \mathcal{Y}_{t+\alpha}, \bar{a}_1 \mathcal{Y}_{t+\alpha-1}, \ldots, \bar{a}_\alpha \mathcal{Y}_{t+1}$. If, however, we can find the $\mathcal{Y}_t$, and the wavelet $\bar{a}_k$, we can make a best prediction.

The Gramian of (20) is

$$(\bar{x}_t, \bar{x}_{t+\tau}) = \left(\sum_{k=0}^{\infty} \bar{a}_k \mathcal{Y}_{t-k}, \sum_{s=0}^{\infty} \bar{a}_s \mathcal{Y}_{t+\tau-s}\right)$$

$$= \sum_{K=0}^{\infty} \sum_{S=0}^{\infty} \bar{a}_K \bar{a}_S \delta_{K,-J+S} G$$

$$= \sum_{S=0}^{\infty} \bar{a}_S \bar{a}_{S+J} G$$

depending only upon G and $\bar{a}_K$, and not explicitly on the y_t. Thus, the power density matrix $\Phi(\omega)$ depends only upon a_K and G. $\Phi(\omega)$ could also be computed directly from the Fourier transform of the wavelet $\bar{a}_K$, were it known, by $\Phi(\omega) = A(\omega) G A^*(\omega)$ (21) Since there is no phase information in (21) about $A(\omega)$, there are an infinite number of wavelets which will give the same power density matrix. However, $\bar{a}_K$ is invertible. Thus, its transform $A(\omega) = \sum_{K=-P}^{\infty} \bar{a}_K G^{\frac{1}{2}} e^{i\omega K}$, must have an inverse transform

$$D(\omega) = \sum_{K=-P}^{\infty} G^{-\frac{1}{2}} \bar{d}_K e^{i\omega K}$$

Apply Jensen's theorem (2) to det $A(\omega) G A^*(\omega)$. If $A(\omega)$ is a well-enough behaved function, the theorem will be applicable. Then

$$\log \det [A(0) G A^*(0)] \leq \frac{1}{2\pi} \int_0^{2\pi} \log \det [A(e^{i\theta}) G A^*(e^{i\theta})] d\theta$$

$$\leq \int_0^{2\pi} \log \det \Phi(e^{i\theta}) d\theta$$

G. is invertible, hence $\det G \neq 0$. If $\det A(z) \neq 0, |z| \leq 1$, we have strict equality and

$$\log \det [A(0) G A^*(0)] = \frac{1}{2\pi} \int_0^{2\pi} \log \det \Phi(e^{i\theta}) d\theta$$

or
$$\det [A(0) G A^*(0)] = e^{\frac{1}{2\pi}\int_0^{2\pi} \log \det \Phi(e^{i\theta})\, d\theta} \qquad (22)$$

This application of Jensen's formula suggests attempting a Kolmogoroff factorization of $\Phi(\omega)$, in the same manner as was done for the scalar case in Part II. A complete analogy would be to expand $\bar{A}^2(z)$ in a matricial power series

$$B_0 + B_1 z + B_2 z^2 + \ldots .$$

where the B_κ are matrices. Unfortunately, equation (22) is a scalar relationship--it relates the determinant of the wavelet with the exponential of the integral of the determinant of $\Phi(\omega)$. There is no simple relationship between a determinant and the elements of the matrix involved. Nor is there any hope that the elements of $\bar{A}$ cannot individually vanish for $|z| \leq 1$. Furthermore, as Wiener and Masani point out, even if we were led to consider an expansion in matrices of the form

$$e^{D_0 + D_1 z + D_2 z^2 + \ldots .}$$

there is a difficult problem of uniqueness, since $e^{A+B} = e^A e^B \neq e^B e^A$ and commutativity has broken down (e^A would be defined by Sylvester's formula, appendix A).

Though it is conceivable that these difficulties may be overcome by proper definition, we are instead led to consider an entirely new approach.

Since $\bar{a}_\kappa = (x_0, \gamma_{-\kappa})$, a knowledge of $\gamma_{-\kappa}$ is

sufficient to get a best fit to the prediction

$$\hat{\bar{X}}_{t+\alpha} = \sum_{\kappa=\alpha}^{\infty} \bar{a}_\kappa \, \vartheta_{t+\alpha-\kappa} \tag{23}$$

Therefore, we consider the inverse problem

$$\vartheta_{-t} = \sum \bar{\lambda}_\ell X_{t-\ell} \tag{24}$$

A solution to (24) determines the ϑ_t and the $\bar{a}_\kappa$ from

$$\bar{a}_\kappa G = (\bar{X}_0, \vartheta_{-\kappa}) = (\bar{X}_0, \sum_{\ell=0}^{\infty} \alpha_\ell \bar{X}_{\kappa-\ell}) = \sum_{\ell=0}^{\infty} (\bar{X}_0, \bar{X}_{\kappa-\ell}) \alpha_\ell^* \tag{25}$$

By stationarity, it would suffice to find ϑ_0 , for (25) becomes $\sum_{\ell=0}^{\infty} (\bar{X}_{\kappa-1}, \bar{X}_{-\ell}) \alpha_\ell^*$.

The key to the method is a Hilbert space theorem and corollary of Von Neumann (1950, p. 55-56) which we will assume applies to $\mathcal{L}_2$ and interpret in terms of a vector space.

Theorem (5)

Let E and F be projection operators onto the subspaces M_1 and M_2 , respectively, of $\mathcal{L}_2$. Let η be the portion of $\mathcal{L}_2$ orthogonal to $M_1 + M_2$ and let Q be the projection onto η . Then

$$\begin{aligned} Q &= I - E - F - E(-F) - F(-E) - E(-F(-E)) - F(-E(-F)) \ldots \\ &= I - E - F + EF + FE - EFE - FEF \ldots . \end{aligned}$$

(This is hardly a rigorous statement of the theorem, but it conveys the meaning.)

The projection operators are just the Hilbert space version of ordinary vector space projections--i.e. dot products. The theorem, if interpreted in vector space says the following:

Consider a vector $\vec{v} = (v_1, v_2, v_3)$ and two other vectors $\mathcal{J} = (\mathcal{J}_1, \mathcal{J}_2, 0)$ and $\rho = (\rho_1, \rho_2, 0)$ ($\mathcal{J}$ not necessarily orthogonal to ρ).

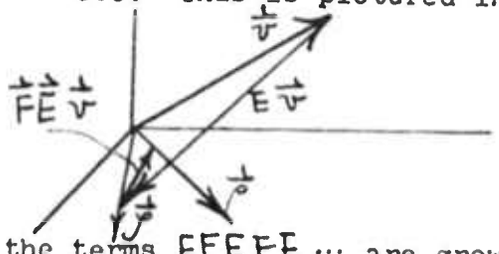
Let the projection of a vector into $\mathcal{J}$ be given by E and the projection onto ρ by F.

Then $E\vec{v} = (\vec{v}\cdot\vec{\mathcal{J}})\vec{\mathcal{J}}$ and $F\vec{v} = (\vec{v}\cdot\vec{\rho})\vec{\rho}$.

Repeating, $FE\vec{v} = (\vec{v}\cdot\vec{\mathcal{J}})(\vec{\mathcal{J}}\cdot\vec{\rho})\vec{\rho}$

$EFE\vec{v} = (\vec{v}\cdot\vec{\mathcal{J}})(\vec{\mathcal{J}}\cdot\vec{\rho})(\vec{\rho}\cdot\vec{\mathcal{J}})\vec{\mathcal{J}} \ldots .$

etc. This is pictured in Figure 2.



(Fig. 2)

Clearly the terms EFEFE ··· are growing smaller and smaller as projections are made, back and forth. In the limit,

$$Q\vec{v} = \vec{v} - E\vec{v} - FE\vec{v} + EFE\vec{v} + \ldots .$$

is just the component of $\vec{v}$ normal to both $\vec{\mathcal{J}}$ and $\vec{\rho}$.

Note that in this space $EF = FE$, and so we need consider only one or the other. In a more general space, the projections do not commute, and both terms are necessary.

If a third vector, λ not in the plane of $\vec{\mathcal{J}}$ and $\vec{\rho}$, and the projection onto it had been considered, the sequence would have gone to zero--a reflection of the 3 dimensions of this space. In a space of higher dimensions, a 3rd operator, or even more, is permissible.

By analogy, the projection operators in $\mathcal{L}_2$ must by the inner products--the Gramians--and the vectors $\vec{\mathcal{J}}$ and $\vec{\rho}$ must correspond to the component one-dimensional time series. If $\overline{X}_t$ has order 3, a sequence of the form

$$I-E-F-G+EF+FE+GF+GE-EFE-\cdots \qquad (26)$$

must be considered, and similarly, for a series of order n.

In applying the Von Neumann theorem explicitly to the prediction problem, restriction will be made to a process of order 2--since from (26) the projection sequence for higher orders is very complex.

Let $X_t = \begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix}$ and let the projection on $X_t^1 = P_1$ and the projection on $X_t^2 = P_2$. Then $P_i = (\ , X_p^i)\, X_p^i$ and the operator applied as $P_i X_t^2 = (X_t^2, X_p^i) X_p^i$.

$P_1 X_t^2 = (X_t^2, X_p^1) X_p^i$ and $P_2 X_t^1 = (X_t^1, X_p^2) X_p^2$ (analogous to $(\vec{v}\cdot\vec{f})\,\vec{f}$). Let $(X_p^i, X_q^i) = \delta_{pq}$ (27) (not a restriction. See Part IV below), and let M_{-t}^1 and M_{-t}^2 denote the pasts of X_t^1 and X_t^2 respectively. Then the portion of X_t, $t = 0, \cdots, -\infty$ normal to $M_{-t}^1 + M_{-t}^2$ must be

$$\bar{X}_t - (\bar{X}_t \mid M_{-t}^1 + M_{-t}^2) = g_0$$

Thus, the innovation g_0 is what will be determined if we can find $M_0 - (M_{-t}^1 + M_{-t}^2)$. Let

$$(X_t^1, X_p^2) = a_{t-p}\ ,\quad (X_t^2, X_p^1) = b_{t-p}$$

then under the condition (27), the projections commute, and

$$g_0^1 = X_0^1 - \sum_m a_m X_{-m}^2 + \sum_m \sum_n b_{m-n} a_n X_{-m}^1 - \sum_m \sum_n \sum_p \sum_q b_{m-n} a_{n-p} b_{p-q} a_q X_{-m}^1$$

$$g_0^2 = X_0^2 - \cdots$$

Factoring $\bar{X}_{-\rho}$ out of each term, this can be written in matrix form as

$$\bar{\mathcal{G}}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \bar{X}_0 - \sum_m \left\{ \begin{bmatrix} 0 & a_m \\ b_m & 0 \end{bmatrix} + \sum \begin{bmatrix} b_{m-n} & a_n \\ 0 & a_{m-n} \end{bmatrix} \sum_{n,\rho} \begin{bmatrix} 0 & a_{m-n}\, b_{n-\rho} a_\rho \\ b_{m-n}\, a_{n-\rho}\, b_n & \end{bmatrix} + \cdots \right\} \bar{X}_{-r} \quad (27a)\ (28)$$

which is seen to be precisely $\mathcal{G}_0 = \sum_m \bar{a}_m \bar{X}_{-m}$ (28) of eqn. (24).

Let $E_m = \begin{bmatrix} 0 & a_m \\ b_m & 0 \end{bmatrix}$. Then (28) becomes

$$\bar{\mathcal{G}}_0 = I\bar{X}_0 - \sum_m E_m + \sum_n E_n E_{m-n} - \sum_n \sum_\rho E_\rho E_{n-\rho} E_{m-n} + \cdots \quad (29)$$

Since $\mathcal{G}_0$ is now explicitly determined, the prediction problem is solved. The expectation of the prediction error is, as in the scalar case,

$$E\left(\hat{X}_{t+\alpha} - \bar{X}_{t+\alpha}\right)^2 = \sum_{K=0}^{S-1} A_K^2$$

4. Computational Procedure

The Wiener-Masani technique has been programmed in Fortran for the IBM 7090 computer at M.I.T., making as great use as possible of existing programs.

There are two main difficulties that present themselves in consideration of a computational procedure.

The first involves the question of convergence of the sequence in eqn. (29). If the "vectors" are parallel, there will be no convergence; if they are nearly so, convergence may be very slow.

The second difficulty occurs in connection with the condition eqn. (27). To show how this is achieved in practice, it is necessary to consider the computing procedure. In actually solving the problem, the method used is a slight variation of the one outlined.

If the power density matrix of $\bar{X}_t$ is $\bar{\Phi}(\omega)$ and if $\mathcal{Y}_t = \sum_m \alpha_m X_{t-m}$, then it can be shown (see Appendix B) that the power density spectrum of $\mathcal{Y}_t$ is

$$\bar{G}(\omega) = \left(\sum_{n=0}^{\infty} a_n e^{ni\omega}\right) \bar{\Phi}(e^{i\omega}) \left(\sum_{n=0}^{\infty} a_n e^{ni\omega}\right)^* \qquad (30)$$

or

$$G(\omega) = \Psi \Phi \Psi^* \qquad (31)$$

But

$$\Phi = AGA^* \qquad (32)$$

Hence

$$G = A^{-1} \Phi A^{*-1} \qquad (33)$$

and from (31)

$$A^{-1} = \Psi \qquad (34)$$

If Ψ can be determined, we can find $A = \Psi^{-1}$ (35)

The problem is then to use the Von Neumann theorem to determine Ψ .

The α_k are given by the matrices of eqn. (29). Hence $\alpha_k = -E_k + \sum_p E_p E_{k-p} + \cdots$, and $\Psi = \sum_m \bar{\alpha}_m e^{mi\theta}$.

In order to insure that condition (27) holds, it is necessary to prefactor the diagonal elements of Φ. That is to say, if Φ_{11} and Φ_{22} are the autopower spectra of X_t^1 and X_t^2, then we find their minimum phase factors such that

$$\Phi_{11}(\omega) = \varphi_1(\omega)\,\bar{\varphi}_1(\omega)$$

$$\Phi_{22}(\omega) = \varphi_2(\omega)\,\bar{\varphi}_2(\omega)$$

The spectral matrix is then

$$\begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix} = \begin{bmatrix} \varphi_1 & 0 \\ 0 & \varphi_2 \end{bmatrix} \begin{bmatrix} 1 & \Phi_{12}/\varphi_1\bar{\varphi}_2 \\ \Phi_{21}/\bar{\varphi}_1\varphi_2 & 1 \end{bmatrix} \begin{bmatrix} \bar{\varphi}_1 & 0 \\ 0 & \bar{\varphi}_2 \end{bmatrix} = \eta\hat{\Phi}\eta^*$$

The problem is worked with $\hat{\Phi}(\omega)$. At the end, the answer is remultiplied by η and η^*. This preliminary factorization is done by either a least squares technique (routine WLLSFP) or the Kolmogoroff technique (FACTOR).

There is some ambiguity about the definition of the wavelet. If one defines $A = \sum_k A_k G^{\frac{1}{2}} e^{ki\theta}$, then of course, $\Phi = AA^*$, a somewhat neater factorization. But with this definition, $A = \Psi^{-1} G$, an explicit knowledge of G is required (which can be obtained from $G = \Psi \Phi \Psi^*$) -- an extra burden on the method.

If a solution is attempted in the time domain alone, one also finds a knowledge of G necessary, for we have $\mathcal{G}_0 = \sum_n \alpha_n X_{-n}$. Then $(X_\kappa, \mathcal{G}_0) = A_\kappa G$ Thus $A_\kappa G = \sum_n \alpha_n^* (X_\kappa, X_{-n})$ and the wavelet is defined only up to the missing factor of $G^{\frac{1}{2}}$.

In practice, of course, one might wish the wavelet inverse to A and the definition $A = \sum_\kappa A_\kappa e^{\kappa i \theta}$ introduces an asymmetry into the relation between A and its inverse.

If we do choose to normalize $\mathcal{G}_\rho$ to $\mathcal{G}_\rho$, $(\mathcal{G}_\rho, \mathcal{G}_\rho) = I$, then we can seek a factorization $X_\kappa = \sum_n C_n \mathcal{G}_{\kappa - n}$ where $C_n = (X_\kappa, \mathcal{G}_n)$, and then the relation between the C_n and the α_κ is completely symmetric.

As a convention, it is perhaps best to require an extended Kolmogoroff normalization $A_0 = I$. Then G and its square roots (see Appendix A) must be positive. Then there is no ambiguity about the factorization.

The matrix nature of the process introduces certain complications into the computations. A flow chart of the complete procedure is appended (fig. 3). The process is still not fully debugged, but the computations carried out thus far indicate that the method is a valid one.

The procedure outlined in the flow chart is carried out by the subroutine WIMAF1. Inputs to the subroutine are two time series, x1 and x2, their lengths,

the lengths of the internal spectral computations, and two cut-off parameters for the Von Neumann projection process--a maximum allowable error in each term, and a maximum number of tries permitted in attempting to reach that error.

Subroutine MACOPS computes the auto- and cross-correlations of the two-time series, (subroutine QXCORR), weights them with a Daniell spectral window (ADANL), and sets up the appropriate power density matrices (ASPECT, XSPECT).

WIMAF1 then calls subroutine FACTOR, which performs the scalar Kolmogoroff factorization of the diagonal members of the spectral matrices. The two resulting wavelets are then taken back into the frequency domain, by COSISP. Subroutine PDDIV performs the division on the spectral matrix, outlined above, that whitens the diagonal elements.

WIMAF1 then drops the constant diagonal terms, and calls COSISP once more taking the now modified spectral matrix back into the time domain. The result is the correlation matrices in the proper form for use in the Von Neumann projection process. The Von Neumann projections are formed by VONEPS (Fig. 4) to within the allowable error set by the input, or until the maximum number of projections permitted is exceeded. The result is the time domain inverse of the variable Ψ . This

is again Fourier transformed by COSISP, and then inverted point-wise in a loop by subroutine COMAIN.

The process is them completed by multiplying Ψ by the wavelets previously taken out by FACTOR. This is performed by COMAML. The result is then the Wold Wavelet.

A program, MAROOT, has also been written for taking the square root of a matrix, if the more complete factorization in terms of G is wanted. This is not, however, currently part of WIMAF1.

The procedure is immediately expandable to orders greater than two with a modification of MACOPS, the setup routine. The key Von Neumann projection subroutine VONEPS, will accept a process of any order. The number of computations to be performed, of course, goes up very rapidly with increasing order.

Since certain processes are best performed with the variables in matrix form (the projections, for example), and some must be done with the variables in linear form (the Fourier transforms), there are a number of conversions into and out of matrix form. This process is rendered considerably earier by a matrix transpose routine, MATRA.

Appendix A

The Square Root of a Matrix.

$G^{\frac{1}{2}} G^{\frac{1}{2}} = G$ is clearly defined when G is a diagonal matrix, if one makes the convention that all roots are positive. Since G is always non-singular Hermitian, it can in principle always be diagonalized.

To obtain the square root when G is not in diagonal form, use may be made of a formula of Sylvester, (Hildebrand, 1952, p. 66).

Let λ_κ, $\kappa = 1, \cdots, n$ be the eigenvalues of an arbitrary Hermitian matrix W. Let $P(x)$ denote a polynomial in x. Then Sylvester's formula states that

$$P(W) = \sum_{\kappa=1}^{n} P(\lambda_\kappa) Z_\kappa(W)$$

where

$$Z_\kappa(W) = \frac{\prod_{r \neq \kappa} (W - \lambda_r I)}{\prod_{r \neq \kappa} (\lambda_\kappa - \lambda_r)}$$

In cases of degeneracy, the limit can be evaluated by l'Hopital's rule.

Let $B = W - I$

Then $W^{\frac{1}{2}} = (I + B)^{\frac{1}{2}}$

The corresponding polynomial is $P(x) = (1+x)^{\frac{1}{2}}$

$$= 1 + \tfrac{1}{2} x - \frac{\frac{1}{2}(\frac{3}{2})}{2!} x^2 + \cdots$$

Assuming the series converges, it can be cut off at x^m, some m, and treated as a finite polynomial.

Then

$$P(B) = 1 + \tfrac{1}{2}B - \frac{\tfrac{1}{2}\left(\tfrac{3}{2}\right)}{2!}B^2 + \cdots$$

$$= P(W-I) = 1 + \tfrac{1}{2}(W-I) - \frac{\tfrac{1}{2}\left(\tfrac{3}{2}\right)}{2!}(W-I)^2 + \cdots$$

$$= W^{\frac{1}{2}}$$

By Sylvester's formula, this becomes

$$P(B) = \sum_{k=1}^{n} P_k \lambda_k \frac{\prod_{r \neq k} (B - \lambda_r I)}{\prod_{r \neq k} (\lambda_k - \lambda_r)}$$

$$= \sum_{k=1}^{n} (1+\lambda_k)^{\frac{1}{2}} \frac{\prod_{r \neq k} (B - \lambda_r I)}{\prod_{r \neq k} (\lambda_k - \lambda_r)}$$

where now the λ_k are the eigenvalues of $W-I$. Hence to evaluate $G^{\frac{1}{2}}$ we need the eigenvalues of $G-I$, a relatively simple process when G is 2x2.

Appendix B

A brief, heuristic proof that $G = \Psi \Phi \Psi^*$

Let $\vartheta_n = \sum_\kappa \alpha_\kappa X_{n-\kappa}$

Then

$$(\vartheta_n , \vartheta_p) = \sum_\ell \sum_\kappa \alpha_\kappa (X_{n-\kappa} , X_{p-\ell}) \alpha_\ell^*$$

$$= \sum_\ell \sum_\kappa \alpha_\kappa R_{n-\kappa-p+\ell} \alpha_\ell^*$$

$$= \sum_\ell \sum_\kappa \alpha_\kappa \left[\int_0^{2\pi} e^{i(n-\kappa-p+\ell)\theta} \Phi(e^{i\theta}) d\theta \right] \alpha_\ell^*$$

$$= \left[\int_0^{2\pi} \sum_\kappa \sum_\ell \bar{\alpha}_\kappa e^{i(n-\kappa-p+\ell)\theta} \Phi(e^{i\theta}) \alpha_\ell^* d\theta \right]$$

$$= \left[\int_0^{2\pi} \left(\sum_\kappa \alpha_\kappa e^{i(n-\kappa)\theta} \right) \Phi(e^{i\theta}) \left(\sum_\ell e^{i(p-\ell)\theta} \alpha_\ell \right)^* d\theta \right]$$

$$= \left[\int_0^{2\pi} e^{i(n-\ell)} \left\{ \left(\sum_\kappa \alpha_\kappa e^{-\kappa i\theta} \right) \Phi(e^{i\theta}) \left(\sum_\ell \alpha_\ell e^{-i\ell\theta} \right)^* \right\} d\theta \right]$$

Hence $(\vartheta_n , \vartheta_p)$ is the Fourier Transform of

$$\left(\sum_\kappa \alpha_\kappa e^{-\kappa i\theta} \right) \Phi(e^{i\theta}) \left(\sum_\ell \alpha_\ell^* e^{-i\ell\theta} \right)$$

But $(\vartheta_n , \vartheta_p) = \delta_{np} G$. Therefore, the Fourier Transform of $(\vartheta_n , \vartheta_p) = G$. Hence by Fourier's theorem,

$$G = \left(\sum_\kappa \alpha_\kappa e^{-\kappa i\theta} \right) \Phi(e^{i\theta}) \left(\sum_\ell \alpha_\ell e^{-i\ell\theta} \right)^*$$

$$= \Psi \Phi \Psi^*$$

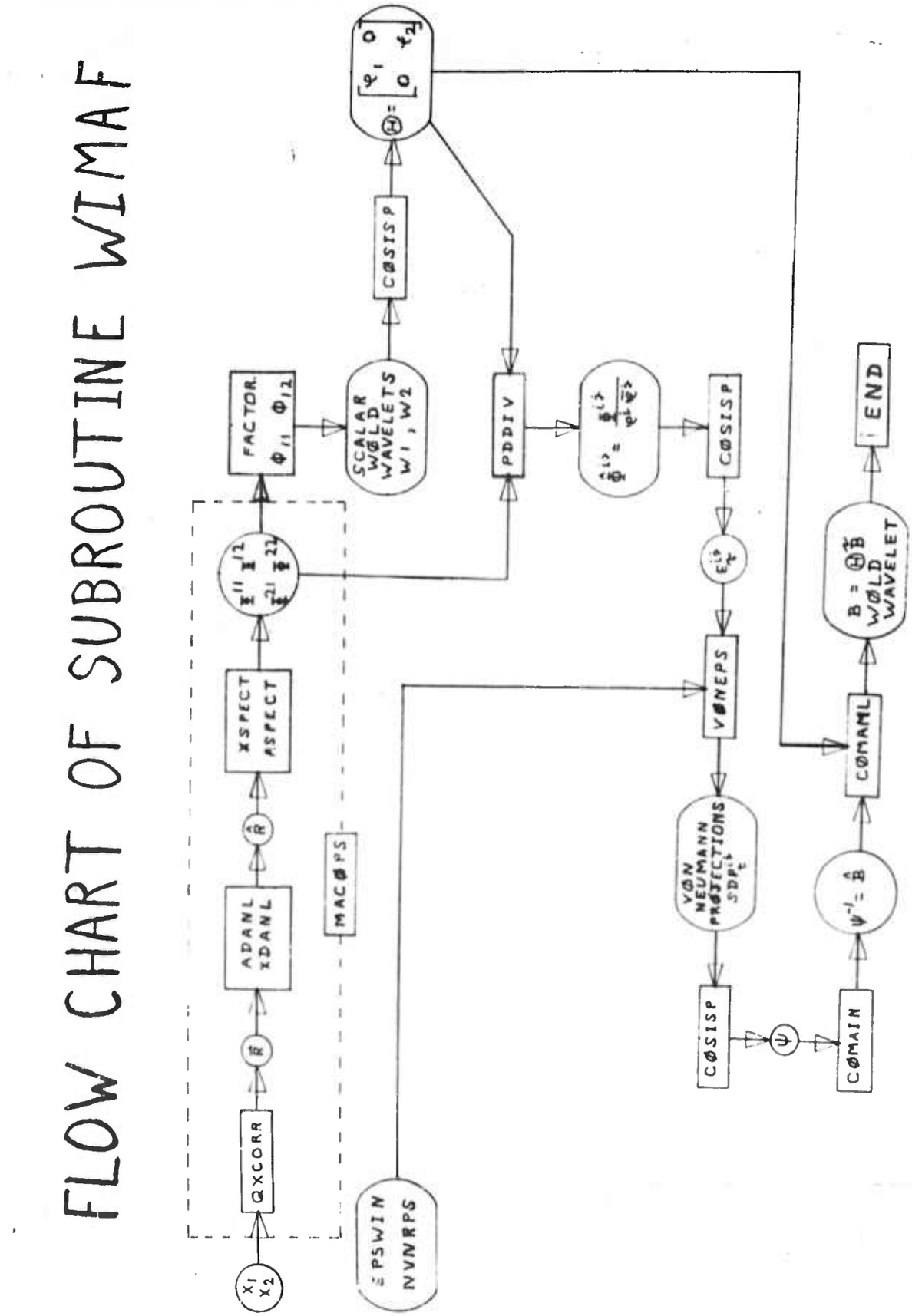

FLOW CHART OF SUBROUTINE WIMAF
QXCORR
ADANL XDANL
XSPECT ASPECT
FACTOR.
SCALAR WØLD WAVELETS W1, W2
CØSISP
PDDIV
CØSISP
VØNEPS
VON NEUMANN PROJECTIONS
CØSISP
CØMAIN
CØMAML
B = WØLD WAVELET
END
MACØPS
ΞPSWIN NVNRPS

Figure 3.

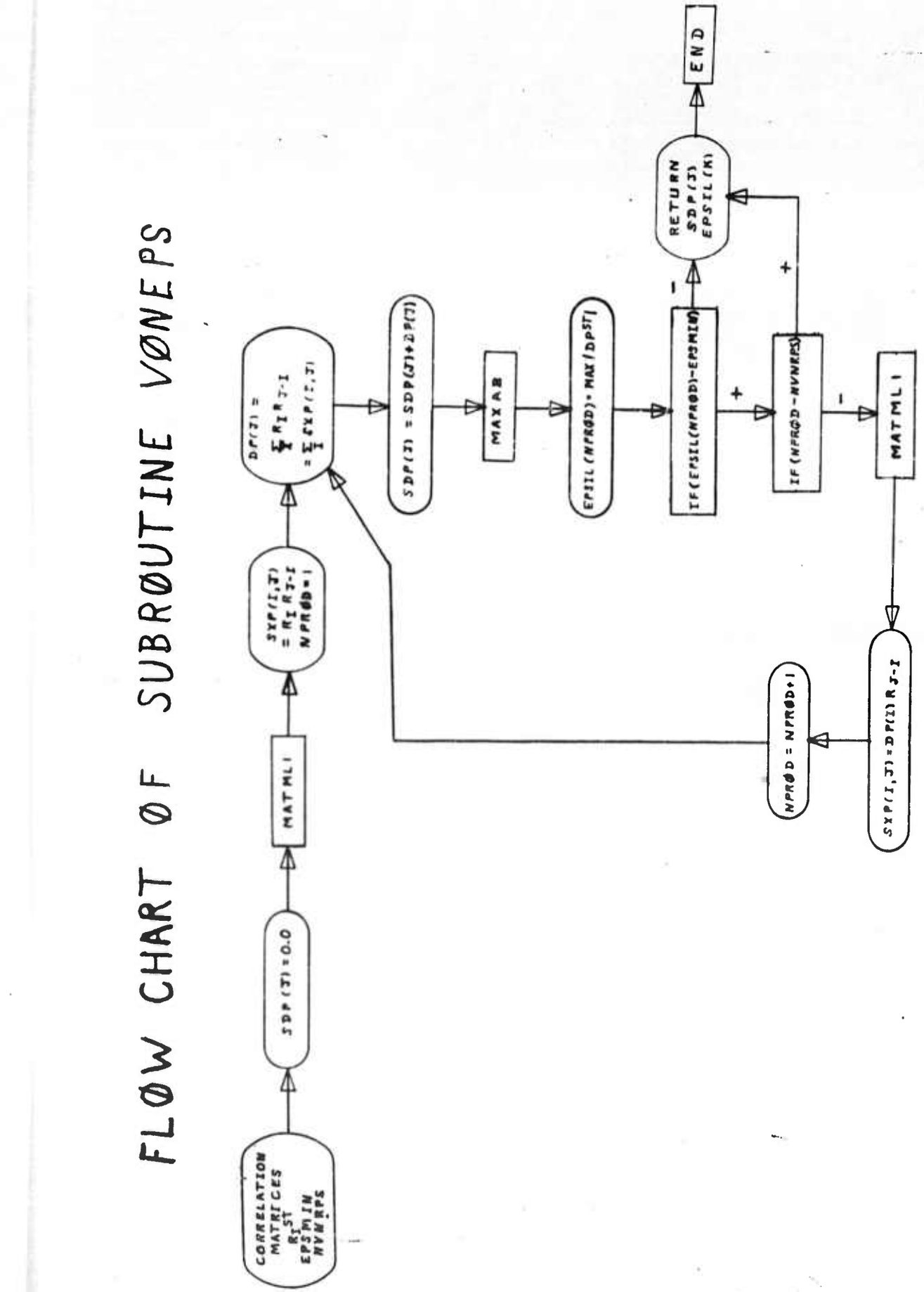

FLOW CHART OF SUBROUTINE VONEPS
CORRELATION MATRICES R_I^ST EPSMIN NVNRPS
SDP(J) = 0.0
MATMLI
SXP(I,J) = R_I R_J-I NPROD=1
DP(J) = Σ R_I R_J-I = Σ SXP(I,J)
SDP(J) = SDP(J)+DP(J)
MAXAB
EPSIL(NPROD) = MAX/DP(J)
IF(EPSIL(NPROD)-EPSMIN)
RETURN SDP(J) EPSIL(K)
END
IF(NPROD-NVNRPS)
MATMLI
NPROD = NPROD+1
SXP(I,J) = DP(I) R_J-I

Figure 4.

REFERENCES

Ahlfors, L. V., 1953, Complex Analysis, McGraw-Hill, New York.

Claerbout, Jon F., 1960, Digital Filters and Applications to Seismic Detection and Discrimination, S. M. Thesis, Massachusetts Institute of Technology.

Doob, J. L., 1953, Stochastic Processes, Wiley, New York.

Halmos, P., 1951, Introduction to Hilbert Space, Chelsea, New York.

Helson, H. and D. Lowdenslager, 1958, Prediction Theory and Fourier Series in Several Variables, Acta. Math., No. 99, p. 165-202.

_____, 1961, Vector Valued Processes, Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2, University of California Press, p. 203-212.

Hildebrand, F. B., 1952, Methods of Applied Mathematics, Prentice-Hall.

Levinson, N., 1950, "The Wiener R.M.S. Error Criterion in Filter Design and Prediction," Appendix B of Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series, M.I.T. Press, p. 129.

Masani, P., 1960, The Prediction Theory of Multivariate Stochastic Processes III, Acta. Math., XXX, No. 104.

Robinson, E. A., 1954, Predictive Decomposition of Time Series with Application to Seismic Exploration, M.I.T., Geophysical Analysis Group Report No. 7.

_____, 1959, An Introduction to Infinitely Many Variates, Charles Griffin & Co., London; and Stechert-Hafner, New York.

_____, 1961, Random Wavelets and Cybernetic Systems, Charles Griffin & Co., London; and Stechert-Hafner, New York.

Simpson, S. M., E. A. Robinson, J. F. Claerbout, J. N. Galbraith, J. Clark, 1961a, Initial Studies on Underground Nuclear Detection With Seismic Data Prepared by a Novel Digitization System; Annual Report No. 1 of Contract AF 19(604)7378 at M.I.T., prepared for Geophysics Research Directorate, Air Force Cambridge Research Laboratory, Office of Aerospace Research, United States Air Force, Bedford, Massachusetts.

Simpson, S. M., E. A. Robinson, J. F. Claerbout, J. N. Galbraith, W. P. Ross, J. Clark, 1961b, Time Series Techniques Applied to Underground Nuclear Detection and Further Digitized Seismic Data; Scientific Report No. 2 of Contract AF 19(604)7378.

Simpson, S. M., E. A. Robinson, J. F. Claerbout, J. Clark, J. N. Galbraith, C. Pan, R. Wiggins, 1962a, Continued Numerical Studies on Underground Nuclear Detection and Further Digitized Seismic Data; Scientific Report No. 3 of Contract AF 19(604)7378.

Simpson, S. M., E. A. Robinson, J. F. Claerbout, J. Clark, J. N. Galbraith, R. J. Greenfield, R. A. Wiggins, 1962b, Magnetic Tape Copies of M.I.T. Geophysics Program Set I (Time Series Programs for the IBM 709, 7090); Scientific Report No. 4 of Contract AF 19(604)7378.

von Neumann, 1950, Functional Operators, Princeton University Press, Volume 2.

Wiener, N. and P. Masani, 1957, The Prediction Theory of Multivariate Stochastic Processes, Acta. Math., No. 98, pp. 111-180.

_____, 1957, The Prediction Theory of Multivariate Stochastic Processes II, Acta. Math. No. 99, pp. 93-137.

Wiener, N., 1950, Extrapolation, Interpolation and Smoothing of Stationary Time Series, M.I.T. Press.

Wiener, N., 1955, On the Factorization of Matrices, Commentarii Math. Helvetici, pp. 97-111.

Yaglom, 1962, An Introduction to the Theory of Stationary Random Functions, translated by R. A. Silverman, Prentice-Hall.